REVIEW AND EVALUATION BULLETINS

Volume 6, Number 3

Performance Appraisal in the School Systems of Ontario

The Honourable Sean Conway, Minister
Bernard J. Shapiro, Deputy Minister

Ministry of Education
Ontario

86-05965/

Performance Appraisal in the School Systems of Ontario

Griffith A. V. Morgan

This research project was funded under contract
by the Ministry of Education, Ontario.
It reflects the views of the author and not
necessarily those of the Ministry.

© QUEEN'S PRINTER FOR ONTARIO, 1986

Order Information:

Publications Sales
The Ontario Institute
for Studies in Education
252 Bloor Street West
Toronto, Ontario M5S 1V6

(416) 926-4707
Will invoice on orders over $30.00.
Other orders must be accompanied by a
cheque or money order payable to
O.I.S.E.

Publications Services
880 Bay Street, 5th Floor
Toronto, Ontario M7A 1N8

(416) 965-6015
(Toll Free) 1-800-268-7540
(Toll Free from area code 807)
Ask operator for Zenith 67200
Order must be accompanied by a cheque
or money order payable to the
Treasurer of Ontario

Contract 0563
ONO 4273

Canadian Cataloguing in Publication Data

Morgan, Griffith A. V.
Performance appraisal in the school systems of
Ontario

(Review and evaluation bulletins, ISSN 0226-7306 ;
v. 6, no. 3)
ISBN 0-7729-1856-2

1. Public schools--Ontario--Evaluation. 2. Teachers--
Ontario--Rating of. 3. Public schools--Ontario--
Employees--Rating of. I. Ontario. Ministry of
Education. II. Title. III. Series.

LB2838.M67 1986 371'.01'09713 C86-090848-8

CONTENTS

		Page
Preface		iv
Section 1	The Reasons for Appraisal	1
Section 2	The Purposes of Appraisal	5
Ssection 3	The Needs of the Organization	8
Section 4	The Nature of Professional Work and Teaching	9
Section 5	Theory and Practice in Appraisal	12
Section 6	Appraisal of Teachers	14
Section 7	Issues and Techniques in Appraisal	20
Section 8	Level and Purposes of Appraisal Systems	24
Section 9	The Field-Site Studies	27
Section 10	The Ontario Study	34
Section 11	The Wider Approach: Organizational and Conceptual Issues	40
Conclusions		60

PREFACE

The Ministry of Education, Ontario commissioned a series of researches into performance appraisal of educators in Ontario.

A comprehensive literature review of research and development in the management, personnel, and educational fields was prepared by Ondrack and Oliver, A Review and Analysis of Performance Appraisal Processes, 1986, volume 1. The same authors carried out a field study of progressive practices in five business or private research enterprises and five colleges of applied arts and technology in Ontario and site studies of seven school systems in the United States, described in volume 2 of the above report.

A large-scale, comprehensive empirical survey of appraisal policies and practices in a representative sample of 30 boards of education (from among 117 boards) in Ontario consisting of a pilot study, a questionnaire study, and eight more detailed case studies was carried out by Lawton, Hickcox, Leithwood, and Musella, the Ontario Institute for Studies in Education, and reported in two volumes of a technical report, Development and Use of Performance Appraisal of Certificated Education Staff in Ontario School Boards, 1986.

A review of the factors bearing on education and appraisal and futurology study and a review of basic concepts in understanding professional work, organization, and appraisal was carried out by a private research-consulting firm, Hay Associates Canada, presented in two volumes, Alternative Futures for Ontario Schools and Some Implications for Performance Appraisal Systems, 1986, by Atkinson, Sussman and Monk, and The Future of Performance Appraisal for Certificated Education Staff in the School Boards of Ontario, 1986, by Hayman and Sussman.

In addition, a non-technical report reviewing the issues in appraisal as demonstrated by literature research and by the field studies was prepared by Ondrack and Oliver A Review and Analysis of Performance Appraisal Processes, 1986, volume 3. A non-technical report, Development and Use of Performance Appraisal of Certificated Education Staff in Ontario School Boards, 1986, volume 4, summarizing the conclusions of the major empirical study was prepared by the authors, Lawton et al. 1986.

It is recommended that, except for archival and detailed research purposes, those interested in gaining a viewpoint on appraisal of educators in relation to the research in education and in management and personnel areas should proceed as follows. They should first study the Hayman and Sussman report, The Future of Performance Appraisal for Certificated Education Staff in the School Boards of Ontario for a cogent outline of the concepts involved in understanding the nature of professional work such as teaching and education appraisal. This report sets the procedures of appraisal clearly within the context of organizational needs and goals. Second, they should review the 1986 non-technical study by Ondrack and Oliver.

This study tends to overgeneralize from its own research base, but the problems of performance appraisal are clearly set fourth in the context of organizational needs and goals and the major literature, concepts, and problems are reviewed. Thirdly, they should review the non-technical report on the survey of appraisal in the school systems of Ontario (Lawton et al. 1986). This last report provides most of the major findings and issues and clearly defines the scope and limitations of its findings. Proceeding in this way, the reader should form a clear idea of what is involved in performance appraisal in the educational system. Then, if necessary, the reader can return to review the detailed technical reports.

Section 1

The Reasons for Appraisal

The reasons advanced for the current interest in improved appraisal methods (Ondrack and Oliver non-technical report 1986) are as follows:

1. changes in the legal environment, particularly in the United States where there is increasing pressure to document compliance with legal and civil rights. It is doubtful whether these changes have had the same impact in Canada;

2. the changing economic environment, with various factors influencing inflation, productivity, investment, and unemployment. In both public and private sectors, there is pressure for more cost effectiveness, efficient use of resources, and demonstration that performance is well managed, with good performance rewarded and weak removed. However, it is estimated that trends in Canada are five to ten years behind those in the United States (Ondrack and Oliver 1986);

3. the social environment comprises changing attitudes to work, with increased desire of employees, particularly professional staff, to participate in decision making, desire for more information about themselves and, in general, changing values emphasizing freedom and individualization in work;

4. demographic trends, with increasing relative proportions of older, senior workers, implications for efficiency and ability to deal with novel situations, and possible decline in motivation. Three potential uses of performance appraisal may be emphasized: human resource planning; motivation of personnel who, because of age or seniority, are on a career plateau; and identification of needs for retraining.

The above factors likely vary in different fields of work. For example, demographic and seniority trends may be presently more marked in the educational system. Additionally, for over a decade, there has been discussion of the need for "accountability" within the educational system: accountability of teachers to students, managers to the school system, and the school system to the political and social community. This trend has been most marked in the United States. However, research or surveys do not indicate that a demand for accountability is the specific reason for the increased interest in performance appraisal in Ontario's educational system.

Coupled with the Ondrack and Oliver review of factors which may affect the trend to appraisal and the demands for relevant, effective, and acceptable appraisal designed to meet such needs, in terms of improved use of staff, improved organizational effectiveness, and improved motivation of staff, there was a "futurological" study in the Atkinson report.

According to this study, conducted among panels or juries of educational experts in Ontario, many dramatic changes in economic, demographic, social, and technological conditions, in Canada in the next 15 years are forecast. These changes include:

- increased emphasis in the short term on job skills such as math, science, and computer literacy;
- reduced emphasis in the short term on liberal and fine arts;
- increased emphasis on family studies and related community courses;
- less enrichment;
- less scope in the programs for remote and sparsely populated areas;
- more emphasis on adult basic literacy;
- increased emphasis on social skills in the medium and long term;
- requirements to reduce emphasis on standard subject content;
- increased emphasis on teaching cognitive and study skills;
- increased emphasis on values.

It is anticipated that:

- the custodial function of the school will decrease;
- there will be more need for cooperative education, bringing teachers into contact with business, industry, and community;
- teachers will be more involved in teaching adults;
- secondary school teachers will need to broaden their subject matter expertise because of decreased enrolment and school sizes;
- fewer specialists will be required;
- teachers may need to focus more on diagnosing learning needs and planning instruction--more instruction will utilize technological aids, e.g., computers;
- teachers will be expected to adjust programs to serve individual needs over a wider range;
- improved personal and counselling skills will be needed by teachers and principals to deal with increased social and personal problems arising from economic and social change;
- principals and educational administrators will need to become more entrepreneurial in order to get the funding, equipment, and personnel needed to keep and develop their programs;
- principals will need means of dealing with their communities to ensure community support;
- principals will be expected to provide professional leadership on curriculum development and implementation and professional management of colleagues, as contrasted with routine administration;
- teachers will need to learn how to delegate instructional tasks to the new technology;

- parents will be increasingly vocal and involved in educational decisions, not only at the classroom but at the school and board level;
- student groups will be increasingly heterogeneous;
- school schedules will need to be more flexible;
- individual school attendance in terms of timetable and an extended and a flexible school day will have to be shaped to meet varying needs.

Some of these predictions, based on a wide range of opinion (Atkinson report) may be contradictory or there may be divergent trends. The known inertia within the school system will make for slow change or attempts to resist or delay until change is inevitable, is mandated politically, or comes explosively.

Performance appraisal systems must serve clear purposes in satisfying the needs of educational organizations as they exist and as they change. Appraisal must perform a significant and just role in: (1) personnel and management decisions such as hiring, promotion, carrying out objectives, and management of staff, (2) the professional development of staff, and (3) the accountability of the educational system to a wide range of "stakeholders".

If one asks, however, what motives are explicit in the present performance appraisal system in Ontario, as exemplified by the recent surveys, it would seem that none of these issues is consistently and clearly offered as a rationale for extended or improved performance appraisal. Currently, much appraisal is done because it is assumed that the methods demonstrate accountability and efficiency or contribute to improved professional development, but mainly "because it is done", because other boards do it, or because what is practised in the United States is considered to be a relevant model to copy. There is no indication in the empirical research study that performance appraisal is linked with on-going management, supervision and monitoring of educational staff, or clearly based on the purposes, goals, and problems of the educational organization.

The warning in the Hayman report is appropriate: that formal appraisal systems should be adopted only when there is a specific need, when existing management and supervisory practices appear inadequate or insufficient, and that appraisal systems must be set in the context of organizational need. There is a fundamental question as to whether this need has been demonstrated in the Ontario survey (Lawton et al. 1986).

The research literature (Ondrack and Oliver 1986) makes comparisons with business and industrial appraisal practices and bases its analysis on organizational and management concepts, which present useful perspectives on educational performance appraisal. Perspectives from other countries would also be useful, as well as perspectives from the most recent research findings in education, e.g., on the problems of measuring effectiveness in specific areas such as teaching reading.

In summary, in view of the various conditions and trends discussed as bearing on the need for performance appraisal in the educational system, there is an urgent need to identify why appraisal systems are adopted and what needs they meet.

The Hayman report indicates that educational systems may rush into appraising the performance of individuals in order to solve real or imagined problems within an organization, whereas the correct solution may be other actions such as reviewing goals or improving communication with the community. The need to design and establish a formal performance appraisal system must be demonstrated.

Section 2

The Purposes of Appraisal

The following section summarizes basic evaluation issues, based on the review of research by Ondrack and Oliver (1986).

The reasons for adopting a formal performance appraisal may vary, but comprise:

- monitoring and control of the performance of personnel;
- administration and management;
- compensation (whether this is related to performance or other factors);
- personnel management: tenure, promotion, transfer, firing;
- human resources and planning to make the best use of staff and resources to meet the organization's plans and goals;
- accountability of the organization to stakeholders such as the community, parents, the government, shareholders.

Appraisals exist to maintain "satisfactory performance in relation to the goals and expectations of the organization". Performance appraisal may be related to compensation, e.g., bonus or merit ratings (increasingly found in industry and private enterprise). Merit and motivation pay is, however, considered difficult to operate in teaching and is, by and large, replaced by considerations such as qualifications, further education, and seniority. These considerations, as seen later, make for a fundamental difference between the design and use of appraisal systems in education and those used by business and industry.

Similarly, appraisal systems may be used for a variety of personnel and management functions such as probation and entry to professional work, retention in post for satisfactory performance, granting of tenure, transfer, or promotion. It has been pointed out (Ondrack and Oliver 1986) that valid and reliable appraisal decisions must be sources of information from which management can make objective, fair decisions and provide effective use of human resources.

As seen later, educational appraisal (field studies, Ondrack and Oliver 1986) serves different purposes and should vary to meet specific needs, as shown by the field studies (Ondrack and Oliver 1986). In some situations, the emphasis may be on ensuring high levels of qualification and competence at entry to the school system and probation. In other situations, appraisal simply guarantees continuing minimum levels of competence. In yet other situations, appraisal exists to ensure the specific competences required by a highly structured system of educational goals and procedures, and yet elsewhere, to bring about the urgent raising of quality of education in a crisis situation where the public has lost confidence.

One appraisal system may be expected to deal with all purposes, from ensuring qualification and competence at the probationary stage or confirming the competences of established teachers to providing for promotion or coping with incompetence (Ondrack and Oliver's multilevel concept). Appraisal systems may have to develop specialized functions to meet varied demands.

In educational appraisal, the emphasis appears to be on maintaining minimum standards of performance and fostering professional development.

An effective appraisal system should help in gathering information on the abilities, knowledge, experience, and potential of all staff with a view to their most effective use to meet the current and future plans of the organization and to ensure the best fit between the individual and the job he is asked to perform (human resource-personnel planning). Information from the Ontario educational system (Lawton et al. 1986) suggests that human resource planning is not a major function of performance appraisal in Ontario.

Organizational accountability would appear to be crucial, i.e., accountability of the members of the system at every level to others for whom they are responsible or to whom they are responsible, e.g., teachers to students, teachers to colleagues and supervisor (manager); supervisors (principals) to staff and to their administrators, administrators to the schools and to their employment board, and board to community. It is open to question whether individual performance appraisals contribute to all these goals and whether the present appraisal systems ensure adequate accountability.

Professional development of individuals is an important outcome of appraisal and is constantly reiterated as a major goal in educational appraisal in Ontario (Lawton et al. 1986). However, professional development is not a privilege or purpose in its own right and it must be closely related to the humane and effective use of human resources by managers and to effective performance within the context of the goals and plans of the organization (Hayman report).

Failure to take this viewpoint results in the feeling gained from reviewing the empirical research in Ontario and the field studies of schools in the United States, that there is much lip-service paid to the need for formative evaluation, free from judgement and with feedback to the individual leading to professional development, but much less evidence that professional development is built into the appraisal system, is provided for, followed up, or even takes place.

It is pointed out by Ondrack and Oliver that the multiple purposes of an appraisal system may be both its greatest strength and its greatest weakness. The major reason for failure of performance appraisal is too many conflicting objectives, and it has even been suggested that systems which combine many functions, from monitoring and judging performance to providing professional development, rarely win both employee and management support.

Perhaps an appraisal system should confine itself to a limited number of objectives or an effort must be made to reconcile and integrate all the purposes which appear relevant. This means careful planning, not always evident in the research literature.

Appraisal of individual or group performance must have criteria and performance standards which are:

Valid: Appraisal methods must actually measure and predict the performances and competences which are important for teaching effectiveness.

Reliable: Assessments and scores must remain reasonably consistent from occasion to occasion, from appraiser to appraiser, or in different situations.

Discriminant: Appraisal judgements can distinguish and maintain significant differences between levels of performance. It should be noted that in classical test theory and in practice, both validity and reliability depend on the capacity of measurements to discriminate between individual performances.

Acceptable: A system may have many admirable technical and organizational merits, but will fail unless it is fair and humane and allows for involvement of the appraised and due recourse in the event of dispute or unjust decision.

Standards of performance must be:

1. based on job and performance, not the individual (though, of course, the individual is assessed by his response);
2. achievable and realistic, in terms of minimum attainment or degree of performance;
3. fully understood;
4. fully accepted by both those appraised and those who appraise;
5. as specific and objectively measurable as is consistent with the work situation, outcomes, goals, and procedures involved;
6. time-related, not vague;
7. written or documented as far as possible, so that there is no ambiguity or bias;
8. subject to change in the face of changing circumstances, goals, and needs.

Section 3

The Needs of the Organization

Much of the research and practical application of performance appraisal in education has taken little or no account of the goals and needs of the organization within which and for whose benefit it takes place. The benefit of the literature review by Ondrack and Oliver (1986) and their broad orientation to the topic of performance appraisal, as well as the important discussion of basic issues in the Hayman and Sussman report has been to introduce perspectives from outside education, from industry, business, and personnel management theory, i.e., the perspective of organization theory.

It is evident that any effective appraisal system must take account of the needs, purposes, goals, and plans of the organization in order to have sense and purpose.

As Ondrack and Oliver 1986 point out in placing appraisal in context, all organizations must attempt to determine the extent to which their employees are performing effectively and carrying out the goals and plans of the organization, whether evaluation is formal or informal. It may be appropriate to systematize and formalize the process though, as the Hayman and Sussman report points out, there must be sound reasons for doing so.

There is a rapidly growing movement (mainly in the United States and in private enterprise) towards accountability for the means by which an organization maintains, develops, and satisfies its goals and plans and makes the most effective and humane use of its human resources.

There is a strong "justificatory" trend in both business and public service, i.e., the need to justify and defend explicitly and legally, on sound and acceptable evidence, organization decisions involving the quality of individual performance. This trend has important legal and civil rights implications.

Performance appraisal must take place in the context of supervisory and management monitoring of performance and the requirement to do this efficiently and acceptably to achieve the goals of the organization, however much emphasis is placed on formative or personal development aims for appraisal. The lack of coherent meaning and purpose in appraisal systems, commented on in the research literature (Ondrack and Oliver 1986, Hayman and Sussman report) is in large part due to lack of clarity of purpose within the organization. This lack of clarity of purpose also emerged in the empirical survey of performance appraisal in Ontario (Lawton et al. 1986).

Both Ondrack and Oliver 1986 and Hayman and Sussman, working from a wide perspective including management and personnel studies, emphasize that the design and use of an appraisal system must be shaped to meet the specific needs of an organization.

Section 4

The Nature of Professional Work and Teaching

The purpose, style, characteristics, and organization of the work or professional activity to be appraised have important implications for the design of the appraisal system. From the point of view of organizational theory and practice, it is evident that it is necessary to have a clear understanding of the "work" to be evaluated. However, there is no clear or agreed theory or practical consensus on what "education" and "teaching" are (Ondrack and Oliver 1986). Research in this area is inconclusive and unsatisfactory. There is more opinion than hard evidence on what constitutes effective teaching, and appraisal practice depends on assumptions, tradition, or convictions based on opinion or experience (Ondrack and Oliver 1986).

There is no clear definition of teaching as a profession.

The need to define the work which is to be appraised has led researchers in organization theory to look at the forms of organization which are labelled "professional" or "professional bureaucracy" such as the legal and medical professions, universities, hospitals, social work agencies, public accounting, and school systems.

Study of its requirements and duties suggest (Mintzberg, in Ondrack and Oliver 1986) that teaching is a species of "professional bureaucracy". A professional bureaucracy is defined as professionals who are duly trained and indoctrinated, then given considerable autonomy and control over their own work. There is normally a requirement for standardization of skills and knowledge, ostensibly rigorous. There is considerable discretion in decisions in performing duties. Professionals tend to identify with professional norms and goals rather than with an organization for which they work. The power of the professional is in expertise, and the output of professional work cannot be readily measured or standardized, in particular by outsiders.

There are important implications for the difficulty of appraisal of professional performance, the kinds of performance that can be evaluated, and the ways in which it can be evaluated.

The Hayman and Sussman report, which is summarized in the final section of this report, presents a useful discussion which takes this analysis further. In a simplified model of varieties of work, it distinguishes between means and ends. Routine work is based on having little choice over means or of ends (D--) as exemplified by a factory-line worker or a bus driver. The next level of choice is where the goals are determined but there is some choice of immediate means (D-+) as in a self-employed taxi driver who has to reach a specified destination but may vary his route. The professional level is best described by the situation in which there is freedom to choose both means (methods, processes, action-plans) and ends (goals, outcomes) or D++ activity.

Teaching may be described as falling within the D-+ or D++ category, i.e., high autonomy, need for freedom to make decisions, responsibility for diagnosis of a problem, planning of intervention, taking appropriate action, and evaluation of outcome. It has been suggested that because of the alleged complexity of teaching, the appraisal of teaching can only take account of the process, i.e., the means and methods used, the plans made, and success in implementing plans, and that measurement of specific outcomes or results, e.g., student achievement, is extremely difficult or inappropriate. The kind of work involved in a D-+ or D++ occupation requires, at some level, the measurement of outcome (results), however difficult or complex. This is discussed further in the final section of this report, in relation to the concepts of professional work advanced in the Hay Associates report, vol. 2.

Appraisal of professional work is complex because, by definition, the process of diagnosis of the problem, planning, developing strategies, carrying out appropriate action and then evaluating the outcome is complex and rests on the decisions of the autonomous professional worker. It is therefore paradoxical that appraisal of teacher performance should rely so heavily on limited "process" approaches such as the observation of brief episodes of classroom teaching by a supervisor such as a principal. The evidence (Ondrack and Oliver field studies, Ontario empirical study, Lawton et al. 1986) is that a major element of performance appraisal for teachers is a narrowly focused classroom observation, with limited concern for other activities which form part of effective teaching.

The claim to being a self-regulated profession with an allegedly high degree of autonomy and freedom has sometimes led to claims that appraisal is irrelevant or inappropriate unless conducted entirely by and for teachers. This appears to be not only impracticable but indefensible, since the profession or professional bureaucracy is employed within a complex educational organization which has to take responsibility for performance and show accountability for its goals, use of staff and resources, and results.

The Hayman and Sussman report examined at greater length in the final section of this report, raises perplexing questions about whether teaching is a profession, in terms of the organizational models of work discussed.

It is evident that a theory or description of work, whether or not one finds the D--, D-+, and D++ model useful, is crucial in order to determine how performance is to be appraised and the goals, criteria, standards, data, and techniques of appraisal. As noted earlier, an adequate analysis of teaching as an activity is still not forthcoming (Ondrack and Oliver 1986).

Listing of "competences" in teaching, as used by some educational appraisal systems in the United States (Ondrack and Oliver field studies) goes some way to providing a ground for a more empirical, objective observation and analysis of effective teaching, but there are

variations in the kinds and specificity of such prescriptive lists, and they still do not determine what is the most appropriate model of the teaching and learning process or demonstrate how competence is related to effective activity in the classroom. Much research remains to be done.

Section 5

Theory and Practice in Appraisal

There are a number of basic issues. Research on performance appraisal has many defects and has been relatively unrewarding (Ondrack and Oliver 1986). It has been among the worst-organized forms of activity (Hayman and Sussman report). Setting up appropriate criteria and standards for appraisal requires a clear and consistent description of the performances to be evaluated. Attempts have been made (Ondrack and Oliver, field-site studies) to develop descriptions of relevant teaching behaviours (competences). There is no commonly accepted description of teacher behaviour, however, establishing performance standards requires a clear, consistent theory of teaching, not yet available. Performance appraisal is known to create anxiety and suspicion, although this may have been the result of poorly planned or executed appraisal systems which were not evidently fair or valid or reliable or did not involve the cooperation of those appraised. Previous experience has led to considerable disenchantment with the entire performance evaluation process (Ondrack and Oliver 1986).

Review of the literature shows that performance appraisal is a complex, controversial subject. There is a wide range of approaches and one approach cannot satisfy all expectations. Recent research on appraisal has not been promising in terms of meeting the practical questions of organizations (Ondrack and Oliver 1986).

This leads to a fundamental difficulty in evaluation of teaching. Although administration, control, personnel-management, and staff development are important outcomes for evaluation, they are not ends in themselves, but only steps in the improvement of instruction and school effectiveness. This improvement, in turn, has as its purpose the production of student learning, achievement, and growth (Ondrack and Oliver 1986). The appraisal of performance is at least twice removed from the final goal, and, as much educational research demonstrates, may not be closely coupled with student learning (see Hoffman 1986, annotated at the end of this report).

There is difficulty in finding "sufficient proof" that teachers are performing at an acceptable level because of lack of consensus on competency and its relationship to actual teaching behaviour and on appropriate criteria and standards for performance. Accountability is difficult or meaningless unless educators know how they are to be accountable and on what evidence and that they can take direct responsibility for the outcomes of learning in view of all the factors which intervene in and affect the teaching process, e.g., student characteristics, curricula, resources, time, and organization. (Ondrack and Oliver 1986, Hoffman 1986).

In many respects, the process of appraisal depends on faith (analogous to our faith in the honesty and effectiveness of the legal system, Ondrack and Oliver 1986). Much depends on the confidence of those participating in the system, in particular, the community which pays for and receives the service. The Hayman and Sussman report points out that the educational system depends on a chain of confidence, each level assuming that the one below it is effective and attempting to avoid direct inspection of outcomes and accountability at any given level. As long as there is confidence, the process continues, but questioning confidence in any part of the process or organization could reveal how little is based on rational or empirical evidence.

Section 6

Appraisal of Teachers

Teacher evaluation covers eight major issues under two main headings (according to Ondrack and Oliver 1986).

Internal administration and supervision:

1. to rationalize the process of making evaluations otherwise based on informal supervisory procedures;
2. to promote professional growth, through giving teachers feedback, direction, and motivation;
3. to provide a valid basis of administrative decisions on promotion, transfer, and possibly compensation;
4. to monitor and control quality of teaching and learning.

Public accountability:

5. to improve instruction and school effectiveness;
6. to enhance student growth and achievement;
7. to improve the general quality of education in a cost-effective way;
8. to provide evidence to educators, teachers, and the public that the goals of teaching are being reached in a fair, effective, and systematic way.

There are three basic criteria for performance: presage, process, and output.

Presage includes the characteristics which the teacher brings to the job, such as personality, professional training, and intelligence. Among those related to teaching effectiveness are scholarship, professional preparation, attitude to teaching, age, and experience. Much research in teacher personality and other personal characteristics has proved fruitless. Presage criteria are more applicable to recruitment and selection, i.e., at entry to the profession and in the probationary period, and of limited use thereafter. However, it is pointed out in the final section of this report, following the Hayman and Sussman report that presage criteria, i.e., strict selection of professionals on the basis of qualifications and skills, form an important part of the appraisal of high-level D++ professionals.

Process evaluation deals with the how of teaching: methods, diagnosis-prescription, action-plans, organization and materials, classroom strategies and skills, teacher-student relationships and motivation, communication, and, possibly, public relations. Process evaluation receives much emphasis in the literature on educational evaluation, and is the major focus of actual appraisal in Ontario (Lawton et al. 1986) and elsewhere (Ondrack and Oliver 1986, field studies of United States school systems). Process evaluation is adopted when it is felt that professionals cannot be adequately evaluated in terms of results, but

only on their on-going professional behaviours. It is also adopted because it is considered too difficult or impolitic to measure outcomes or results such as student learning or tested achievement.

The problem with process appraisal is that it emphasizes what the teacher does, rather than how and what the student learns. It runs into the problem of defining what teacher behaviours and competences should be, as already mentioned. There has been no demonstration of a complete and comprehensive set of teaching competences which are shown to be directly related to teacher effectiveness in terms of student learning. Process appraisal must also assume that some given set of professional behaviours and techniques is more appropriate or valid than another, i.e., that the "method is right". If professional performance actually validates itself in outcomes, e.g., the improvement of a client, then the preferred techniques tend to be validated and reinforced. However, there is the danger that the approved technique, e.g., class discussion, open education, systematic didactic teaching, become ends in themselves.

Business and private enterprise, and, to some degree, community colleges, are much less likely to adopt process appraisal. In enterprises devoted to and measured by outcomes, whether in the market, in research and development, or in client change and satisfaction, the methods of obtaining results are less important than demonstration of outcomes. There may be no agreement that there are "best methods".

Appraisal by outcomes or results is more characteristic of private enterprise (Ondrack and Oliver 1986, field studies) and, to some extent, community colleges because of the expectations of students who pay for and evaluate their courses and of the employers of graduates from the colleges. Results are assessed in terms of various forms of student learning or change of behaviour.

Appraisal by results has been usually avoided in schools and disliked by teachers (see Ontario survey, Lawton et al. 1986) because of the alleged difficulty of measurement and, in particular, the problems of demonstrating a direct relationship between student gain and the teaching behaviour of a teacher or the program in a school (Ondrack and Oliver 1986, Hoffman 1986). It has been claimed that what a student brings to school in terms of ability, motivation, prior knowledge, and home and peer influence is very powerful in determining learning. Student achievement may reflect only part of the total objectives for which the school is accountable and it is difficult to adjust final standing in terms of standing at the beginning of instruction. It has almost been suggested that such influences are so powerful as to reduce the effect of the teacher very considerably.

If this is believed, then the implication is that teaching has a minor influence on learning. If this is really so, then not only is the relevance of appraisal of teaching in question, but also the purpose of teaching itself. In such an event, it would be more appropriate to attend to factors in the student and his/her environment which can be modified

or to place emphasis on programs, teaching resources, and organization as they affect individual learning rather than pursue the appraisal of teacher performance.

A danger of using outcomes, particularly in the form of tested student achievement, is "teaching to the test" or making the goals of the examination the goals of teaching, or teaching to minimum standards to ensure high student achievement by the majority, or teaching lower-level skills or knowledge which can be more readily taught and tested. This has occurred historically.

All these considerations are valid, but one gets the impression that the difficulties are made to appear insurmountable. If there are problems in using results, in particular short-term student achievement, then the organization of goals and assessment should be modified to deal with these, for example, through the use of more imaginative and better-controlled forms of assessment. This means that the evaluation of student achievement must in itself form an important topic of research and praxis, as a component of appraisal within the educational system.

Little attempt appears to have been made to develop systems of establishing standards, not merely by tests and examinations, but by juries of teachers or consultants to set up and moderate assessment of achievement, as in the British Certificate of Secondary Education, or to combine goal setting and test achievement as in the Montgomery County, Maryland, U.S.A. school system (noted in Ondrack and Oliver, field studies), or to calibrate the performance of schools in order to control and adjust the variation between schools and between students.

It is clear, however, that at some level, the educational system must produce evidence of outcomes from the educational process, not simply descriptions of how teaching is performed or how far teaching styles conform to what is believed to be the correct or most effective or in tune with current ideologies.

There is evidence (Ondrack and Oliver, field studies) that some school systems in the United States are attempting to deal with this problem. It is of interest that in one school (Montgomery County, Maryland) which is attempting to use student evaluations of teachers, there has already been a decade of development in the evaluation of student achievement.

It has been proposed that a hybrid system, in which both process and outcomes are combined, might overcome the limitations of each (Ondrack and Oliver 1986). Recent educational research on effectiveness of teaching reading, for example, has developed a process-product approach (Hoffman 1986). It is asserted that combining process and outcome is more valid than either method alone. The problem is one of complexity and cost. However, the full and appropriate appraisal of D++ level professional work demands review of all aspects of process (expertise) as well as product.

A variety of evaluation techniques are available. The most obvious and most used is appraisal by a supervisor, usually the teacher's principal or the principal's superintendent (Ontario survey, Lawton et al. 1986). This method requires that the supervisor is fair, unbiased, and skilled and has received training for his task.

A perennial problem is the reluctance or inability of a supervisor to act as both judge (evaluating) and counsel (advising, providing guidance for personal or professional development) or as coach to instruct and remedy. This problem can be overcome by suitable organization and training or division of roles. Much depends on the effectiveness, fairness, validity, and, in particular, the acceptability of the whole appraisal system.

However, the evaluation of teachers is still based mainly on a limited period of classroom observation, (Ontario survey, Lawton et al. 1986 and several of the American school systems studies in the Ondrack and Oliver field study). Classroom observation is a narrow approach, dealing with only a small portion of what should be a teacher's full professional repertoire: diagnosis, planning, choice of teaching strategies and materials, evaluation of learning, and relationship with students and colleagues.

Observation is subject to personal biases and restrictions on the adequate sampling of classroom behaviours, even when detailed guidelines are used. In any event, as the Ontario survey (Lawton et al. 1986) shows, rather little time and usually on only one occasion is given to this central focus of appraisal of a teacher by a principal. Classroom observation and its implications for evaluation of "correct" or "acceptable" procedures by the teacher contradicts the assumption of the teacher's being a D-+ or D++ level professional worker.

A second possibility in appraisal is self-evaluation, for which there is a strong claim within a professional setting in which there is allegedly high autonomy and responsibility and difficulty in external appraisal. However, the research does not demonstrate that self-evaluation in itself is sufficient. Within the demands of an organization such as the educational system, there must be other forms of supervision and appraisal to ensure accountability (Ondrack and Oliver 1986). Self-evaluation is rather rare in teacher appraisal (Ontario survey, Lawton, et al. 1986, Ondrack and Oliver field studies of American schools and Ontario community colleges).

Within a professional setting, the claim of peer appraisal is a cogent one. In professions such as medicine and the law, peer appraisal of an overt or covert kind is crucial. It is also an article of faith in another teaching system, the universities. There is no doubt that peer evaluation is valuable. Peers should be more aware of the realistic demands of the work situation and its resources than supervisors. However, problems occur when peers compete for the same rewards or resources. There are opportunities for bias and subjectivity and a tendency to appraise all colleagues as acceptable. Teachers oppose peer evaluation (Ondrack and Oliver 1986), a finding that was confirmed by the Ontario survey (Lawton et al. 1986). The existence of a strong teachers' union also discourages the acceptance of peer evaluation, since a union tends to foster "universalism" and "equality" as

desirable norms, and peer evaluation may be seen as creating an atmosphere of comparison and competition.

Appraisal by the clients (students) would appear to be relevant for a professional service. It is evident that within a free-enterprise, free-choice system, and even within a professional bureaucracy such as medicine, the client's evaluation and choice of professional has weight. Another teaching system, the universities, has accepted student evaluations of courses and teachers. A problem with student evaluation is validity and consistency, and the effects of various factors such as popularity, student bias, student performance, size of class, and type and demands of course. Hence there is considerable controversy on the use of student evaluations. On the other hand, research indicates a significant though modest relationship between student achievement and student ratings of courses.

It is clear that student evaluations cannot be the sole or major aspect of appraisal. However, they appear to have a contribution to make. At least one American school system (Montgomery County, Maryland) used student ratings in a carefully organized way and appeared to find them beneficial. It is to be noted that this school system also used student achievement results and has a decade of experience using tested student achievement as an appraisal measure.

Relevant questions are: Why does the educational system makes so little use of appraisal from its clients, students, and/or consumers such as parents? Why do teachers (Ontario survey, Lawton et al. 1986) appear so lukewarm or negative to this source of information which is, after all, used daily by the teacher and school in monitoring and adjusting effective teaching behaviour?

In business, industry, and private enterprise there has been increased use of goal setting or "management by objectives". In this approach, the individual and supervisor discuss and develop mutually agreed-upon goals and plans to reach these goals. Evaluation consists of checking how effectively this program of goals has been achieved or how far action has been successfully taken to implement the plans to reach such goals. It has been claimed that this approach improves individual performance, motivation, and feedback, enabling the worker to adjust performance or goals.

Problems with the management by objectives approach include the tendency to set goals which can be reached, rather than those which are more complex or challenging, and the setting of personal goals which do not necessarily relate to the plans and goals of the organization. Management by objectives cannot be left entirely to the individual worker and supervisor but must be monitored for agreement with overall organization goals, plans, and accountability.

In teaching, there are several advantages to a form of appaisal such as management by objectives, or a similar technique, "clinical supervision". It helps to clarify roles and responsibilities, encouraging cooperation between teacher and supervisor, encouraging meaningful input from the person appraised, and acknowledging special circumstances involved. Goal setting tends to motivate performance and self-evaluation is encouraged. However, the approach is time consuming. It is difficult for teachers in a class setting to identify and formulate realistic long-term goals. Special skills and training are required. The relatively long time frame needed for assessment of student gain reduces the effectiveness of feedback and modification of performance. The approach is difficult to intergrate into an administrative and control system although it is very useful for personal development.

Both the Ontario survey and the American field studies indicated that there are elements of objective setting in teacher appraisal, insofar as the supervisor provides a preconference and postconference discussion of what has been observed or the information gathered in individual teacher appraisal. However, it is clear that this is a minor element, and the fact that there is little evidence that there is organized provision of individual professional development or follow-up of recommendations for this (Ontario survey, Lawton et al. 1986) means that there is little effective management by objectives, at least within the formal appraisal system for teachers.

Other possible methods include appraisal by a group of persons rather than the individual principal and/or superintendent who usually carries out observation and interviewing (Ontario survey, Lawton et al. 1986). This method could be extended to visits by an external group of unbiased, expert personnel, e.g., consultants, master teachers. There is evidence for limited use of this approach in Ontario as it is viewed as an "inspectorial" system, a development about which there are reservations. However, there is good potential in an appraisal team for reducing bias and increasing validity, range, and depth in appraisal, and for bringing in experience and expertise from outside the school, the family of schools, or the board system.

In the event that the Ontario Ministry of Education decides to play a positive role in creating guidelines or specifying desirable or minimum standards or criteria for the design or implementation of performance appraisal, the use of appraisal teams for performance as well as program appraisal should receive immediate attention.

Educational administrators already feel that appraisal makes considerable demands on time, if correctly conducted with reasonable frequency. In spite of the problems of cost and complexity, it would be sensible to develop appraisal teams that have time and resources allotted to them, rather than assuming that principals and superintendents will continue to find the time and resources to conduct appraisals with the frequency and depth required, as well as dealing with their full-time administrative, monitoring, advisory, or planning duties.

Section 7

Issues and Techniques in Appraisal

Much has been made of the distinction, in the evaluation literature, between "formative" and "summative" evaluation. In general terms, formative evaluation refers to on-going assessment and monitoring which can lead to changes in method, materials, techniques, and professional development and which therefore is adaptive, not final. Summative, on the other hand, refers to final judgements on effectiveness of performance in relation to supervisory and organizational goals.

It is clear from the Ontario survey (Lawton et al. 1986) that formative is used in a number of different ways, such as informal vs. formal, nonadministrative vs. administrative, supervision vs. evaluation. From this, it is evident that the term has an ambiguous and vague meaning, with emotive and political overtones.

Formative is contrasted with summative, and it is implied that formative evaluation is more closely related to the purposes of professional improvement. In this sense, it appears to be more acceptable than an approach which emphasizes the monitoring and evaluation of performance in the light of its effectiveness in meeting organizational goals. In fact, the contrast is false. Both formative and summative evaluations are necessary; the one method should contribute to the effectiveness of the other.

Another contrast which emerges in the literature and the field studies is the emphasis on evaluation for management and organizational goals and evaluation which contributes to personal and professional development. The educational literature and the field studies imply that professional development is a major aim and justification of appraisal, if not the major purpose. There is no contradiction between the two aims. Personal and professional development is a sensible, logical outcome of performance appraisal, if aimed at improving individual performance, ensuring a better match to organization goals, and contributing to effective planning and use of human resources.

Personal and professional development can exist only within a framework of organizational plans and goals, i.e., within a broader perspective of management and personnel appraisal. To be effective, or simply to be honest, the organization must ensure that professional development is not only available, but also that it is used and that it is consistent with the goals of the organization. Furthermore, it must be tailored to both the individual's and the organization's needs and plans. There must be effective feedback to the individual and effective follow-up to ensure that professional development has been carried out, whether this is by the organization or by the individual acting according to the code of his profession or as required by his self-regulating profession (as in social work, medicine, law).

In the Ontario survey (Lawton et al. 1986), although professional development was claimed as a fundamental purpose of appraisal, there was little provision for ensuring that professional development took place or was followed up or evaluated. The evidence suggests that much professional development for teachers, principals, and superintendents has broad or diffuse aims, and there is little evidence of the tailoring of professional development precisely to the needs of the individual or of the programs to which he contributes.

The literature review (Ondrack and Oliver 1986) deals with a range of specific techniques used in performance appraisal. There is little purpose in reviewing these in detail, and the reader is referred to the original references.

In summary, a variety of techniques have different strengths and weaknesses when the criteria of objectivity, validity, reliability, lack of bias, acceptability, and economy of time and resources are applied. No particular technique can be considered the best because a technique only makes sense and is effective if it meets the needs of an appraisal system and organization in given circumstances at a given period of time. Much research on instruments is cast in the psychological and educational framework as if an instrument or technique had validity in its own right rather than within a given situation for specific purposes.

There are, however, some general issues. Much of the discussion of instruments concerns individual rating, ranking, critical incident, or testing techniques likely to be applied by one person (supervisor). Little consideration is given to the possibility and problems of appraisal by a group (e.g., an appraisal team) and the relevance of instruments to that situation.

Similarly, it is generally assumed that the appraisal is of an individual. The issue of comparing and ranking within a group is mentioned, but not developed. Comparisons between individuals are not the basic concern of appraisal. They are viewed negatively by those being appraised, e.g., teachers. Union policies and traditions also discourage comparing workers. Nevertheless, there may be situations in which it makes sense to compare teachers with respect to knowledge, skills, or contribution to a joint enterprise. The situation might arise in considering the use of appraisal for selection or promotion to a specific job. It might be useful to compare staff in terms of their contribution to developing or implementing a program.

It would be of beneficial to discuss a number of issues omitted by the research review. These issues include the calibration of individual appraisers to make them more objective and effective, and review of the use of appraisal teams to ensure calibration of appraisals across individuals and schools, as well as representing a range and variety of viewpoints during the appraisal.

Little attention has been given to the issues of using evaluation by peers or by clients (students), or to guidelines for self-appraisal, or to the means of linking self-appraisals with monitoring the performance of the individual for conformity with the plans and goals of the organization. There is one example, in the school-site studies (Ondrack and Oliver 1986) which cites Montgomery County, Maryland as an example of how information from students and parents can be used in appraisal.

If outcomes (results) are to make their proper contribution to appraisal and accountability, there is need for research into how this information (e.g., students' achievement, test and examination results) can best be used. If there is a problem in comparing student learning because of variations in student characteristics, background, school, or program variables, there must be an effort to find how to control these effects in looking at the classroom and school. If student achievement is to be used as a measure, it would seem essential to calibrate it so that reasonable comparisons may be made between the performance of individuals, classes, or schools, and in the differences between initial standing and final status in learning. Calibration methods have been developed to deal with problems of educational evaluation in other educational jurisdictions, e.g., Great Britain.

Again, attention is needed to what might be called measurement of the quality of the product of learning, i.e., to imaginative methods and criteria for assessing student achievement. Other jurisdictions, e.g., Great Britain, have developed alternative ways of assessing student achievement outside formal examinations and tests by using expert teacher panels or juries to replace such tests or to moderate and calibrate both curriculum and examinations.

The effective use of program evaluation requires attention, since program evaluation essentially assesses not only curriculum and methodology, plans, and goals, but also indirectly assesses the contribution of teacher performance in achieving educational goals within the program. Program evaluation is a major instrument for assessing the effectiveness of a system of performance appraisal.

Apart from a review of attempts to create objective forms of judgement linked to specific tasks or "critical incidents" in performance (e.g., the BARS technique, created by and confined to industry), there is little evidence from the Ondrack and Oliver literature review of imaginative, innovative effort in developing techniques for educational appraisal. This contrasts with the evidence from areas of recent educational research such as that into the effective teaching of reading, in which highly detailed and objective forms of observation and analysis of classroom and program variables have been developed (Hoffman 1986).

Considerable stress is laid on the need for objectivity, validity, reliability, and acceptability of an appraisal system. This is merely lip-service. Rarely, if ever, do appraisal systems produce empirical evidence on these issues, i.e., evaluate themselves (Ondrack and Oliver 1986, Ontario survey, Lawton et al. 1986). With one-person appraisers

(principal or supervisor) as is normal, it is difficult to see how reliability (consistency from time to time, occasion to occasion, person to person) or objectivity can be measured. In an appraisal team, this would be technically possible.

Validity is an empty concept if it cannot be measured in operational terms. There is rarely hard evidence for the effectiveness of any appraisal system in terms of distinguishing between teaching behaviour and competence or the relationship between teacher competence and student learning. There is, for example, only indirect and limited evidence on "effectiveness" of performance appraisal systems in Ontario. This is user opinion and judgement, which is "acceptability" rather than proved effectiveness (Ontario survey, Lawton et al. 1986).

Section 8

Level and Purposes of Appraisal Systems

The research places emphasis on the merits of a multipurpose system of appraisal, one which can meet a variety of goals or needs (Ondrack and Oliver 1986), but it is also concluded that having too many purposes contributes to the lack of effectiveness of an appraisal system. It is clear that an appraisal system must differentiate its purposes or provide a variety of levels to meet different needs.

For new entrants or probationers, the appraisal must be relatively frequent, intensive, and detailed, to ensure adequate qualification and competence on the job, as in American school systems which rigorously certify their new staff although they have received certification from colleges of education and the state. By contrast, the appraisal of established teachers may focus on "maintenance" of minimum levels of performance periodically, on the assumption that the system is basically healthy. A different focus may be needed to detect and select staff for advanced training, promotion, or change of role. Specific measures may be needed for special cases such as to "salvage" or proceed to dismiss teachers who are incompetent. Different situations and demands will shape the kind and purpose of appraisal. The literature and the field studies of American schools (Ondrack and Oliver 1986) illustrate this.

In one system, with a rapid change of personnel and the need to recruit a large number of well-qualified staff, the obvious response was to set up a procedure for qualifying and certifying new entrances rigorously in terms of tests and proved performance. In another system, with questions about the quality of education and adoption of a highly structured "masterly" approach in teaching and evaluating students, the emphasis was appropriately on process, i.e., competence in the specific skills and procedures required by the system. In a stable system, with no concerns about quality of goals or teaching, the emphasis was on maintenance of a steady state by monitoring periodically to ensure minimum performance.

It is clear that appraisal systems cannot be independent of the organization's needs, purposes, goals, and specific needs in a given situation and given time. To survey systems as if all served a common purpose (Ontario survey, Lawton et al. 1986) is unrewarding, as is the idea that there may be one "best" appraisal system. Appraisal systems are designed, developed, and implemented in various ways, and remain static or change.

Appraisal systems may be initiated in different ways. A basic distinction is between "top-down" implementation, in which senior management decide on the issues and goals and devise appraisal systems to meet these purposes, and "collaborative" implementation in which

those who are to be appraised contribute to the design and development of the system. Different systems have different sources. The top-down approach is more likely in business and private enterprise with distinct management hierarchies, strong control of goals and plans, and emphasis on results or response to market or clients. It is also found to a greater degree in the community colleges (Ondrack and Oliver, field studies 1986).

In education, the tendency is to devise appraisal systems with the collaboration of those appraised. This ensures acceptability, which may be more crucial than validity or reliability. Further, when dealing with a strong union, as in teaching, it is crucial to ensure cooperation and to negotiate acceptability. However, it might be noted that in a significant proportion of cases in Ontario, appraisal systems were board policy or designed by senior educational staff, with little evidence of collaboration by teaching staff. The conclusion from research is that the most successful systems are likely to be based on a collaborative approach (e.g., examples from the American school field study). However, there is no compelling evidence for this view.

There is a difference between designing and maintaining an appraisal system. The research shows that much effort goes into design and little, if any, into maintaining and up-dating the system. An appraisal system must justify itself and provide evidence of its effectiveness and its cost benefit. The research (Ondrack and Oliver 1986) shows that there is rarely any attempt to monitor and evaluate a system. Once established, it is assumed to be working well. There is little effort to up-date and modify the system in response to experience and evaluation. Indeed, so much effort can go into designing a system that little is left to monitor and up-date it. There is a marked dependency on faith. Related to this is the finding that much effort may be made to train supervisors when a system is first established (American schools field studies, Ondrack and Oliver 1986), but increasingly less effort is devoted to training new supervisors or up-dating. This finding emphasizes once more the need to develop program evaluation to provide evidence of the effectiveness of appraisal systems.

Appraisal systems must decide whether they are to deal only with teachers or the full range of educational staff. Appraisal systems which claim to deal with the full range of the educational profession must address the issue of whether the same techniques and goals can be used to evaluate educational personnel from probationer teacher to senior administrators.

The research literature offers little evidence on the appraisal of educational staff other than teachers and principals. Indeed, the Ontario survey (Lawton et al. 1986) provides the major source of evidence. It is evident that appraisal of principals in most jurisdictions involves a variety of roles and functions, emphasizing goal setting, planning, ability to supervise, administration, and leadership. This is very different from the current appraisal of teachers. It is also noted (Ontario survey, Lawton et al. 1986) that the appraisal of a principal in effect means appraisal of the school in which he operates, i.e., a species of program evaluation.

The evidence of the Ontario survey is clear. The higher in the educational hierarchy, the less organized, purposeful, and effective appraisal appears to be. Appraisal of senior administrators such as superintendents is diffuse and haphazard, though it does acknowledge the need to focus on capacity to administrate, to meet organizational goals, to plan, and to offer educational leadership. The appraisal of directors of education is without clear form or purpose, and essentially incidental and anecdotal. So, although the appraisal system varies in function and purpose to deal with different levels of the educational organization, it does so without explicit planning. Only teachers and principals are subject to systematic performance appraisal.

The response of senior administrators indicates that their job or performance goals are not clearly defined. There is no organized, comprehensive provision for following up on appraisal, for professional development, and for checking on the outcomes of professional development (Ontario survey, Lawton et al. 1986).

This situation is most unsatisfactory. It makes little sense to examine the performance of individual teachers and principals in detail when senior personnel who are responsible for direction, planning, and leadership and who are accountable to the public are not subject to comprehensive, systematic appraisal using the best management and orgnizational principles. Quis custodiet ipsos custodes?

Section 9

The Field-Site Studies

The field-site studies carried out by Ondrack and Oliver are among the most important sources of evidence, since they permit comparisons between education and non-educational professional forms of appraisal, comparisons within education settings, and comparisons between the field studies and the major Ontario survey.

Five organizations which represented research, development, or other professional work in a business or private enterprise setting in Ontario were selected as representing advanced practice. Comparisons were made with five community colleges in Ontario which also led the field in appraisal practices. Both systems were compared with seven school systems in the United States which represented a wide range of characteristics and approaches to appraisal. This evidence was augmented by discussions with Educational Testing Service, Princeton, for many years the major research and development authority on tests, assessments, and evaluation in the United States

Comparisons were intended to explore the similarities and differences in organization, goals, and uses of appraisal or to draw analogies between systems.

Six issues were identified: the differences in appraisal methods in different forms of organization; reasons for the differences; the different functions of appraisal in different systems or levels of systems; varying approaches to gathering data and using criteria; links between management and appraisal; and emphasis on process or outcomes as appraisal criteria.

Nine criteria were explored ranging over: design and maintenance of the appraisal system; criteria and standards; training of appraisers; evaluation of effectiveness of the appraisal system; the handling of special cases such as incompetence and dismissal; and future trends.

There were clear differences between business and education organizations. Business emphasized improving organizational purposes and efficiency and personnel resources, and improving and upgrading professional service. Education emphasized the mandatory requirements for appraisal and improving quality of professional service and personnel administration.

Business emphasized career organization and use of appraisal for administrative and personnel decisions involving reward for performance. Education assumed steady tenure, and appraisal is not used on any scale to determine rates of compensation, promotion, or dismissal. Business made use of output criteria, i.e., results. There was more use of goal

setting and planning by individuals and groups and management by objectives was used more. Education focused on appraisal of process. Business emphasized measurement of performance on the job, whereas education relied on maintaining minimum standards and seniority. There was more emphasis on personal responsibility in business, whereas in education it was felt that the teacher cannot be held fully accountable. Business relied on internal accountability, since top management and market or client satisfaction provided the external criteria, whereas education has many more stakeholders and is open to public and political scrutiny. Business appraisal procedures were generally more varied and inclusive, whereas educational appraisals relied considerably on narrow-focus procedures such as classroom observation. The approach was more top-down in business and more collaborative in the educational systems. However, a substantial number of the American school systems used a top-down approach, especially when confronted with an educational crisis or political pressures. It was noted earlier that there is a substantial measure of top-down approach in the Ontario educational system (Ontario survey, Lawton et al. 1986).

The Ontario community colleges fell somewhere between business and the schools. The colleges have to satisfy both internal clients (students) who pay directly for tuition and external clients (industry, commerce) who employ their students. Colleges made more use of student evaluations than did schools. There was some use of management by objectives, but there was considerable reliance on classroom observation for appraisal by supervisors, as in the school system.

Comparisons between business and colleges revealed the following differences and similarities. Business placed more emphasis on using (high) minimum standards to ensure effective performance related to the goals of the organization and also used appraisal to determine compensation, merit or bonus payments, and promotion or dismissal. This contrasted sharply with practice in the colleges, which relied on selection of staff and monitoring of performance during probation, since staff could not be assumed to have basic teaching qualifications, but acceptance of qualifications "on testimonial" after that stage and reliance on maintaining acceptable minimum standards after final appointment.

Both business and colleges used appraisal to identify a need for personal and professional development but, paradoxically, business was more likely to ensure that there was individual training whereas colleges relied on ad hoc "aggregate" in-service provision or upgraded teaching staff routinely. Although much emphasis was placed by colleges on professional development, appraisal was not used to identify performance problems. In both business and colleges, there was reliance on monitoring by supervisors to deal with problems and incompetence.

Business had moved away from reliance on observation and checklists and used action-plans and results. There was a sharp difference between business and colleges in evaluating action-plans. Business saw the action-plan as a stage leading to the final result, so that the action-plan had to specify in detail what was to be done and outline the sequence and

timing to arrive at the final goals. In the college setting the statement of action-plan was treated as a result in itself, e.g., it was satisfactory to claim that an action-plan was "to attend two appropriate conferences on a given technique" without specifying what results for teaching would follow.

Job descriptions were used in both business and colleges, but were usually general in nature, though more specific in the business setting. The tendency was to specify minimum levels of professional or technical qualification, on the assumption that this was a valid measure of competence and future performance. In this sense, professionals through their qualifications determined the specification of their job.

Whereas colleges relied on testimonials for minimum credentials, business specified minimum professional standards and used an on-going check on competence through the yearly goal setting and planning sessions. Baselines for standards of performance and for the related competences required were thus modified periodically in business.

Neither business nor colleges, despite the claimed emphasis on professional development, ensured that there was systematic follow-up to make sure that professional development was undertaken or that it was effective. Astonishingly, neither used the satisfactory completion of professional development or upgrading as a component of appraisal in the next annual cycle of appraisal.

In terms of the cultural expectations of the organization, there were no explicit codes demanding particular kinds of behaviour or contribution to the organization. However, in three businesses of the five, there were implicit expectations that professionals would fulfil certain obligations, ranging from producing research which was acceptable to a critical academic audience to being active in the community or taking on society memberships. In two colleges, it was clearly expected that staff would be active in making contacts with the industry or commerce served by the colleges. The researchers point out that unless the cultural expectations of organizations are made clear and explicit, they can cause difficulties and confusion, i.e., they are subjective and unpredictable aspects of personal appraisal, particularly when potential for promotion is being considered.

In comparisons across business, colleges, and schools, the following trends were found. All organizations tend to have only general job descriptions. All have some form of certification for professionals, but reliance on qualifications and professional status is much less in business, which relies on reassessment of on-going performance and results. Colleges and schools used planning and objective setting in various degrees at classroom, department, principal's level, etc., but rarely specified specific results to be obtained. The use of management by objectives and the use of appraisal to contribute to the plans and goals of the organization were more prominent in business, as contrasted with the emphasis on professional development in school and college. Nevertheless, as noted before, professional development was taken more seriously by business. Reports by subordinates on supervisors were

rarely used. There was restricted use and no particular pattern of self-report. The colleges relied heavily on reports by immediate clients, i.e., student course evaluations and direct representation by students to supervisors. A major contrast was the reliance on process appraisal in colleges and schools and results in business.

Turning now to the American school system systems. These covered a variety of size, geographic location, and educational philosophy. They covered those in which appraisal systems were directed by the state, those with both state and local contributions, and those which were local to a school system or district.

All school systems endorsed aims of appraisal in three major areas: improved student learning, improved quality of education, and professional development. There was also agreement on the need for feedback of appraisal to the teacher and for accountability by the system, but in no instance was appraisal linked with compensation or personnel decisions (tenure, promotion, dismissal).

Job descriptions did not generally provide specific guidelines for individual appraisal, though two of the five systems did have more specific job descriptions. It was found that a job description would lead to dispute as to what a teacher was or was not accountable for.

All systems had minimum standards of professional qualification, but it is of interest that whereas five accepted academic accreditation from the training institutions, up-dating their decision on staff after the probationary period, three systems insisted on local certification in addition to state certification, indicating a lack of confidence in the capacity of the teacher training and certification agencies. Insistence on control of certification was found when there was a marked increase of recruitment of new staff in an expanding system or a nearly complete lack of confidence in the quality of education in the system. As discussed later, if teaching is a professional activity in the fullest D++ sense, then selection for high professional qualifications at entry is crucial.

Generally, appraisal was used to maintain the system, to ensure minmum standards in the belief that the system was basically healthy. This is generally the case in Ontario, also (Ontario survey, Lawton et al. 1986). Appraisal was not linked with compensation, merit, or bonus pay, though teachers might be paid for extra duties. A similar situation is found in Ontario.

It was claimed that appraisal identified training and professional development needs, but there was little evidence of individualized professional development or systematic follow-up on this. This contradiction was also clearly indicated in the Ontario survey. Although American school systems had master teachers and advanced teaching certificates, none used appraisal to detect or guide potential for moving towards this advanced training. This is odd, since the same systems used appraisal to get teachers to work towards required

certification and a number of systems used appraisal to encourage professional development beyond initial certification.

The main method of appraisal was classroom observation and a formative approach, i.e., discussion and feedback in a goal free, non-judgemental setting aimed at professional development. The link between formative and summative evaluations was not close. Recommendations for professional development were not consistently followed through. The situation is similar in Ontario.

Lists of teacher competences of various length and complexity were used by all systems. In Ontario, the use of detailed checklists of this kind varied considerably. A minority of systems used action-plans by teachers as evidence in appraisal. A more standardized format was usually employed for standard, formal stages such as certification or probation but not for the appraisal of established teachers. There was restricted use of management by objectives by teachers, mainly among experienced staff.

In relying on classroom observation, all systems attempted to identify effective teaching behaviours. A majority rated teachers on specific competences related to the individual school system or school. This practice is different from Ontario's. A problem in using teacher competences, however, was the lack of a comprehensive theory of what is meant by education and teaching. Hence, there was reliance on local theories developed through experience and discussion and based on checklists of competences.

There were few references to self-report as a method of appraisal, but more in proportion to the number of school systems than in Ontario. There was also little use of student evaluation, but again relatively more than in Ontario. It was rare to have subordinates evaluate their supervisors.

Student achievement was not used in direct evaluation but in two cases was an element in program evaluation, i.e., the effectiveness of the whole program and staff contributing to it. Four systems, however, measured student achievement but did not include it formally in a final appraisal. There was, however, relatively more use of student appraisal than in Ontario.

There was no formal policy on feedback of appraisal results to the individual teacher. Indeed, three systems claimed to be so large as to make this individual response difficult, which casts some doubt on the effectiveness of the appraisal. In the Ontario survey, feedback through a postappraisal conference with the principal was relatively well established even if the conference was brief, so in this respect at least, Ontario seems to make more effective use of feedback to teachers.

Business, colleges, and school systems were compared. The review showed that for all these systems, the provision of preappraisal conference and planning was weak. In the Ontario survey, the provision of a preconference with clearly defined objectives was a very positive

issue for teachers. Without a preconference, it is difficult to establish a baseline and to set goals for appraisal, so that much of the benefit of postappraisal discussion is forfeited. Business organizations overcame this difficulty by using the annual planning and goal-setting conference as the baseline for later appraisal.

The frequency of formal appraisal ranged from annually to every three years. In this respect, there was more frequent formal checking of performance than in Ontario, where the modal frequency is a three-year interval. In all systems, the supervisor, usually the principal, was the primary communicator, appraiser, data gatherer, and deliverer of feedback. Most systems, especially the relatively new ones, had detailed manuals and guidelines for appraisal. Most drew on administrative support, consulting services, meetings with senior management, and meetings of supervisors concerned with appraisal. However, the training and monitoring of new supervisors dropped off steadily, meaning that the system, once put into operation, was left to look after itself and was not adequately maintained.

The literature review points out that the issue of training for appraisal is crucial. Supervisors need training in many aspects of appraisal if they are to provide valid, reliable, objective, and acceptable appraisals. Teachers also need training in setting goals, developing criteria of performance, and preparing for self-evaluation and professional development plans. Overall, little specific attention was given in the systems studies to the appraisal of supervisors themselves as supervisors (e.g., principals and educational administrators in the education service).

All organizations gave close attention, though not necessarily within the formal performance appraisal, to special cases and obviously inadequate performance. All systems provided for review of problem cases and discussion of plans for salvaging personnel in difficulty. It is clear that, because of legal and civil-rights issues, all organizations must pay careful attention to due process in establishing incompetence and dismissal, and it appeared that there was strong reluctance to proceed as far as this formally, as compared with finding alternative ways of dealing with problem cases. It may be recalled that in Ontario, a minute proportion of teachers are subject to review for incompetence each year and even fewer are dismissed. This situation raises fundamental questions.

In the American schools, the most serious problems were with older teachers who had passed their probation long before rigorous performance appraisal had been introduced. The Ontario survey also reports a high proportion of teachers currently in mid-career, especially at the secondary school level, and the Hay Associates reports pointed out the problems which can ensue from this accumulation of staff whose training lies far in the past and who remain in post because of seniority. The discussion points out that the organization risks tolerating mediocrity unless there is a conscientious effort to deal with this issue.

It is disturbing that in no instance was there any evidence of monitoring the effectiveness of the appraisal system. School systems rely on faith for the validity of their

appraisals. To quote: "Many of these systems are operating with appraisal systems that take it as an article of faith that the particular certification criteria, list of competences, goals, etc., are actively producing results." In the Ontario survey (Lawton et al. 1986) there was no evidence of research- or practice-based evaluation of appraisal and the research study itself relied on approximate and subjective judgements of satisfaction. To quote, in conclusion: "The evaluation of an appraisal system is essentially a test of program evaluation; that is, an assessment of the outputs and consequences of a program in comparison with the goals."

Section 10

The Ontario Study

Many of the issues raised by the survey of performance appraisal in the school boards of Ontario (Lawton et al. 1986) have already been related to state-of-the-art practices in American schools and other comparable educational and professional contexts (Ondrack and Oliver field studies 1986).

The Ontario survey (Lawton et al. 1986) presents a massive amount of empirical data about practices in the mid-1980s. A basic weakness of the study is that, despite its use of a model of the development of a performance appraisal system, it does not place appraisal in the context of organizational and management theory. It is limited by its educational focus and techniques. Its findings are essentially descriptive, with no clear implications for policy development, except possibly negative ones. Most of the data are based on questionnaire findings which therefore reflect opinion and judgement, not direct fact or process.

A crucial finding is summed up in the statement on page 6 of the non-technical report: "One of the general findings ... is that while a lot is known about what makes an effective set of appraisal policies and procedures, school systems in Ontario have not implemented their policies consistently in many jurisdictions."

Although a strength of the study is that it concerned itself with the appraisal of educators at all levels of the school system, performance appraisal is most common and comprehensive at the teacher and classroom level and least common and effective at the senior administrative level.

Much emphasis was placed by the informants on the importance of formative appraisal but, as noted earlier, the term "formative" was used with several interpretations, e.g., administrative vs. non-administrative use of appraisal, formal development, classroom observation vs. comprehensive review. As commented by the Hay Associates report, the distinction between formative and summative is not of primary importance and obscures the real issues involved.

There were basic problems in this study, therefore, with the variety and ambiguity of the terms used by informants in describing appraisal systems.

As expected, two major purposes of appraisal were mentioned: appraisal to bring about change in performance and appraisal to enable judgement to be made about the value of performance.

The major appraiser of teachers was normally one person, the principal, with the support of a superintendent in a minority of cases. The main appraisal procedure was observation of classroom performance, preceded in a substantial minority of cases (36 per cent) by a preconference and followed up by a postappraisal conference in only 24 per cent of cases, for an average of 21 minutes discussion. Time spent on appraisal of an individual was a day or less. In about half the cases, documentation such as the teacher's plans was reviewed, though only 27 per cent of teachers felt they should be. This is a very narrow base of appraisal, as also pointed out in the field studies and review of literature. Teachers themselves would prefer to have more interviews, as compared with classroom observation. Self-evaluation was hardly ever used. There was effectively no use of client-student evaluations. Appraisal by colleagues was not used, and teachers reported negative attitudes to this issue.

Standards and criteria were external in the sense that 38 per cent were board policy, 48 per cent were set by the supervisor, and only 10 per cent were collaborative. The reader is referred back to the previous discussion of the contrast between top-down and collaborative approaches to designing an appraisal systems. Clearly, in Ontario, it is top-down.

However, a high proportion of teachers (76 per cent) claimed that they provided information for their appraisal and a very high portion (88 per cent) considered they should. Discussion of setting of objectives was referred to in substantial minority of cases (38 per cent). There was high compliance with appraisal policies.

As noted earlier, there was a major emphasis on professional development as an aim of appraisal. However, the actual outcome is somewhat devastating. Only 14 per cent of teachers reported developing plans for personal professional development as a result of their appraisal and even fewer, 7 per cent, reported that follow-up was monitored. To quote: "The general feeling is that not much is done ... in reporting and follow-up."

A large majority of teachers (84.1 per cent) considered there was no improvement or only a small amount following appraisal. Obviously, concern should be felt that so many teachers believe this. What is also disturbing is that superintendents had the reverse perception, 86 per cent seeing modest to substantial improvement. With this degree of disagreement, there is lack of communication and coherence in the system.

The appraisal of principals, mainly in the hands of individual superintendents, was more complex. Apart from the supervisor's appraisal, there was self-appraisal and review of goal setting, planning, administration, and management. The basic assumption is that the principal is evaluated on how well the school is operating.

Fifty per cent of school boards required a postappraisal conference between principal and appraiser (more than for teachers), and principals almost universally reported having such a conference. However, as with teachers, only a minority (25 per cent) of principals reported developing a plan for professional improvement as a result of appraisal and only 15 per cent

were monitored for follow-up. As with teachers, there was little perception of improvement as a result of appraisal. The majority of teachers saw little or no change in a principal as a result of appraisal, and about 70 per cent of the principals themselves saw no improvement. Once again, the perceptions of the senior administrators were quite the reverse, 87.5 per cent seeing improvement in their principals. Once again, one must ask whose judgement is correct and how marked is the lack of communication. This basic issue is summed up in the comment of the authors: "In general, we cannot find much evidence to show improvement in performance that would come anywhere near equality with the amount of effort put into evaluation."

As noted, there was a high degree of perceived compliance with appraisal policy. About half the principals placed high priority on implementing evaluation policy, as did three-quarters of superintendents. A high proportion of teachers (71 per cent) believed that the purpose of their last evaluation had been clearly communicated.

A devastating finding, however, is that when asked for the primary purpose of appraisal, a majority of teachers (76 per cent) and about the same proportion of principals believed it was simply to comply with board policy.

It may be recalled that American school systems had formal appraisal every one to three years. In Ontario, the typical interval for formal appraisal is once in three years for teachers and once in five for principals. Whereas a full, formal appraisal done frequently is cumbersome and time consuming, nevertheless a review once in three years (with the evidence of lack of follow-up, monitoring, or perception of improvement) suggests that there is little continuous and comprehensive control or management of teacher performance. This finding alone suggests that real-life and real-time management of the educational system is likely to be done by supervision and monitoring, not formal appraisal.

If there are concerns over the formal appraisal of teachers and principals, there must be even more over the performance appraisal of the senior administrators in school systems who are responsible for monitoring, managing, planning, administrating, and providing leadership. Superintendent appraisal is not systematic in most school boards. Only 17 per cent reported follow-up appraisal. Interviewing by the director of education was the most common form of appraisal, with some self-evaluation and opportunity to discuss management and planning objectives. There are different appraisals at different levels of the system. In-service courses for senior administrators were not perceived as particularly useful or directly related to individual needs. There was little input from colleagues, teachers, or principals for whom the superintendents was responsible, though superintendents themselves felt this would be useful.

Appraisal of directors of education, when it took place, was even more fragmentary, haphazard, and lacking in clear goals and procedures. There was a distinct feeling among directors themselves that they would welcome more comprehensive and effective appraisal. Often, appraisal simply consisted of meeting with trustees, who are themselves not well placed

in terms of time and expertise to conduct full, effective appraisals. In summary: "There is little evidence to show appraisal of directors is done effectively." The most important component in the school system in terms of management, planning, leadership and possible innovation is the one most neglected. "The higher in the hierarchy, the less concern there is about the purpose of appraisal."

No specific attention is given to the certification or probation stages at which teaching staff are selected in terms of professional qualifications and capabilities, as would be demanded if the teaching profession is indeed highly trained, with specific knowledge and skills, as required by a D++ model of professional service, or the educational system is endeavouring to ensure the highest quality in its staff. Clearly, if the probationary and certification stages are receiving their due attention, this is through informal modes of supervision.

At the other end of the scale, the appraisal systems in Ontario, as in the comparative American school systems, do not focus specifically on special review cases. The survey revealed that, in a typical Ontario school board, only one or two teachers are placed under review each year and a principal in one to two years. Negative ratings of superintendents are nearly non-existent. Termination of a teacher for unsatisfactory performance is even rarer, occurring about once a year in a typical school board. Termination of a principal occurs once in every five to ten years. Appeals and grievance occur about once in five to fifteen years.

The authors themselves indicate the difficulty in interpreting these facts. They can be taken to mean a low rate of incompetence and failure in the school system and also as justifying performance appraisal in setting the stage for dealing with problems ahead of time. However, there is no evidence either way to support this interpretation. Given what is intuitively probable about the wide range of professional competence in teachers, it seems more likely that the formal performance appraisal system is either not detecting or not acknowledging below-standard performance. Coupled with the facts about the perceived failure of the appraisal system to ensure improvement in teaching staff, fundamental questions about the purpose and effectiveness of appraisal are raised.

There is clearly a need to guarantee a range and variety of skills in appraisers, even though there appears to be a reasonably high level of teacher acceptance of appraisal. An attempt by the authors to estimate the effectiveness of appraisal is acknowledged to be approximate and arbitrary. Essentially, it is an estimate of the satisfaction of teachers and principals with various aspects of appraisal based on low though significant correlations with these factors.

Not unexpectedly, the significant positive relationships for teachers were: preconferencing; setting of clear objectives in evaluation; clear statement of purposes ahead of time (related to satisfaction with report forms, perceived fairness of evaluation, skill of

evaluator, serious attitude towards appraisal and achievement of personal goals); use of classroom observation; specific notes and suggestions about classroom observations; and the existence of a final conference after appraisal. Although students were rarely involved, there was a positive attitude to using their information and although development of teaching plans was not reported often, it had a strong relationship to the perceived effectiveness of the appraisal.

Involvement of teachers and principals in the design and implementation of the system had a strong positive relationship to the direct and indirect measures of effectiveness, which is an interesting comment on the actual top-down development of most systems.

For principals, the important variables were: satisfaction with the report form; fairness and skill of appraiser; and fairness and seriousness of the appraisal. For the effectiveness variables, the components were the effectiveness of the system as judged by director; improvement in principal's performance reported by teachers and principals themselves; and improvements as reported by directors.

The important issues for principals were: preconferencing (as for teachers); the clear statement of purposes in appraisal; involvement of teachers in giving information; use of an appraisal team; time spent on appraisal and its relationship to how seriously the appraisal was viewed; and the development of a plan for professional improvement.

No factual, controlled evidence for the effectiveness of the appraisal systems exists. The evidence of the survey points to the recent adoption of many systems and their being left to run with no systematic provision for revision of policies and procedures, as also noted in the literature and the comparative study of American school systems.

In summary, "While it is clear that most school boards in the province have formal appraisal systems, it is not so clear that they are being used with any particular effectiveness. Certainly, if the objective is to improve instruction, the recipients of the process do not sense any particular improvement. ... While there is a tremendous amount of effort put into evaluation by supervisors in many boards, we could not really say that the results are used to any great effect." These reserved, if not negative, findings raise fundamental questions about the purposes of performance appraisal systems and whether they solve the actual problems confronting the school systems, as the Hay Associates report points out.

Compared with the American school systems, the majority of which had policy at the state level, the appraisal systems in Ontario are entirely local, with no central guidelines. Variety and choice are necessary, and each system must meet its own circumstances, but in this case variation appears to have been purchased at the cost of a balkanizing of effort and knowledge.

This study constitutes a major effort in terms of data gathering and the attempt to describe the design and use of performance appraisal systems in Ontario. It goes beyond other sources of evidence in dealing with senior educational staff, though omitting important educational staff such as consultants. However, the effort is descriptive rather than analytic. The evidence is based on report and opinion, via questionnaire, with one limited set of case studies. There is no direct evidence through observation of the working of the systems. It accepts the assumptions of the system, that it should deal with process and does not inquire into results. A basic weakness of the survey is that it needed to be set in the framework of organization theory and a wider perspective on the issues of appraisal in professional settings. The survey is consistent, for better or worse, with findings from a parallel set of school systems in the United States.

Clearly, what should not occur, in view of the conclusions of the literature and field-study reviews and the present findings, is an attempt to patch up the present appraisal system or to adopt bits and pieces from elsewhere. To say the least, in the words of the authors: "Some thoughtful changes will have to be made in most board policies and practices."

What the study accomplished is essentially a house cleaning or a demolition type of research, which closes down a whole area of research as not likely to be useful and clears the ground for discussion of major issues and an alternative approach.

Section 11

The Wider Approach: Organizational and Conceptual Ideas

This section is essentially a review and summary of the contribution of the Hayman and Sussman report to the basic concepts and issues in performance appraisal.

Performance appraisal must be viewed in the context of the purposes and goals of an organization, its management, and monitoring of work. This history of formal performance appraisal is less than inspiring and in many respects, few organizational activities are so poorly done. The field is underconceptualized and underdeveloped. The Ontario survey (Lawton et al.) illustrates the weakness of a purely empirical approach bound to educational assumptions alone and indicates a plethora of systems, criteria, and standards.

The Hayman and Sussman report makes the point that if performance appraisal operates independently of the fundamental standards and operations of the organization (as seems to have happened in Ontario) then it is a hollow ritual. However, evaluation methods can make an organization more effective by changing its goals, structures, and operations.

Performance appraisal was initially developed without too much regard for the nature of organization and management. It occurred first in highly organized work in large, static, hierarchical, industrial organizations and the application of its methods have depended on assumptions about the relevance of transferring these techniques to other, different contexts such as the professions and education.

What is needed, before any performance appraisal is undertaken, is clarification of organization goals. Headings outlined by the Hay Associates reports are: (1) describing the research question; (2) looking at the force driving efforts to formalize appraisal; (3) the four basic responsibilities of management in organizations for: effective planning, organizing, staffing, and supervising; (4) the nature of teaching and the characteristics of the educational system; (5) re-examination of performance appraisal practices; and (6) recommendations.

Under the third heading, the point is made that performance management strategies must precede effective appraisal and must derive from clearly articulated, consistent, appropriate conceptions of the nature of the work being done and managed. Before evaluating workers in their organizations, there must be an explicit and appropriate choice of how they are managed and assigned responsibility, i.e., the nature of work in the educational system. The model proposed, referred to as the D--, D-+ and D++ patterns, is that work differs to the degree to which workers are required and expected to use discretion in performance, in terms of both the goals and the means used to reach the goals.

It is pointed out that the appraisal of performance is a pervasive, everyday activity. Organizations do things and judgements must be made continually on how well those things are done. Clients judge services and employees judge the nature of their work assignments. The question is if, when, and how such on-going judgements and monitoring are to become formalized. Supervisors engaged in management and monitoring usually believe they know who their effective staff are. They may use formal appraisal not because it facilitates performance but simply because it is required of them. This is clearly the case in Ontario, as far as teachers and principals are concerned.

The formal systematizing of appraisal usually goes together with the need to become accountable and to set up safeguards, i.e., to become objective, valid, reliable, etc. The danger is in the assumption that such an appraisal system can measure and quantify without passing judgement. In the last analysis, judgements must always be made. It is questionable whether judgements can be divorced from appraisal.

The identification of effective appraisal must be made with the broad understanding of all the strategies and processes employed and, in particular, the match between the appraisal and the nature of the work appraised. Identification of the purpose of an appraisal system must take account of the specific context in which it occurs. There can be no universal and general type of appraisal. This is illustrated by the field studies (Ondrack and Oliver 1986) which identified specific circumstances and purposes which determined the nature and emphasis of the appraisal adopted by the American school systems in question. The relationship between appraisal purposes and practices is therefore close and critical. Organizations exist for purposes, means to ends, not for themselves, though it sometimes seems that schools and education claim to exist in their own right.

As discussed earlier, the call for formal appraisal appears to follow various types of pressure for accountability (Ondrack and Oliver literature review and field studies 1986, Atkinson et al.). However, the information required to provide the public with proved accountability is not the same as that needed to describe and facilitate teachers' performances.

In repeating that appraisal must be linked with organization goals, structures, and strategies, it is evident that organizations differ in many ways. They develop positions in relation to their environment (society, clients, regulations, stakeholders, competition). They have histories, basic values, and systems of reward and punishment whether formal or informal. They assign tasks to members and must devise methods for stimulating and integrating effort. Organizations have specific cultures (as illustrated by the field studies of Ondrack and Oliver). Hence, one must have knowledge of the specific purposes and culture of an organization such as a school system, a school board, or a school.

Prior to any formal appraisal, organizations manage and monitor performance. A basic aim is to establish goals such as: What is our business? and What should our business be?

In this connection, the goals formulated by the educational system, such as the Ontario Ministry of Education and boards of education, are usually very general. Even curriculum guidelines which are basic developmental and management documents do not give specific direction on procedures. It is difficult for the education system to say, "What is our business?" much less "What should our business be?"

Personnel appraisal is the task of assessing the match between actual and expected performance and is effective only to the degree that expectations (e.g., job descriptions, plans, goals) are clearly articulated in terms of the purposes and circumstances of the organization. In this connection, the Ontario survey (Lawton et al. 1986) noted weakness in the linkage between job description and performance appraisal. The appraisal of principals, for example, may detail lists of criteria for performance but not link well with actual job description or performance. Appraisals may be inappropriately used to encourage activities which are not relevant or not fully accepted by all parties to appraisal.

Without clear criteria, managers and supervisors cannot rationally develop objectives and create a common effort. In particular, professional service organizations such as the education system must think through priorities, select targets, set standards of achievement, and develop appraisal which is as far as possible self-controlled by those participating.

It is necessary to audit objectives periodically to ensure that they remain appropriate and to eliminate obsolete or unproductive activities. So school boards and schools need sunset clauses on their programs and activities, or they may persist in doing what they do because they have always done so.

The Ontario Ministry of Education has not developed adequate standards for performance effectiveness which could apply as minimum requirements across the educational system.

Making work productive and workers into achievers is the major responsibility of management. However, this is not necessarily done by instituting formal appraisal systems.

The Hay Associates reports cites ten examples or case studies of possible difficulties which could arise in the educational setting, involving students, teachers, administrators, and taxpayers. From consideration of such cases, it is clear that the managers of the educational system need to decide what the real problems are and how they are to be dealt with. Indeed, they may often be dealt with otherwise than by appraisal.

A fundamental finding is that the whole system may need evaluation, not only its teachers.

Much effort has been devoted, in the research literature and the Ontario appraisal systems, to formal appraisal. However, the crucial processes are the supervision, monitoring, and management of work in such a way as to make it consistent with the purposes, practices, and standards of the organization.

The most efficient way to control and motivate performance is likely to be self-regulation and self-monitoring. It has been concluded that the more judgements that have to be made external to the worker, the less efficient they are likely to be. However, self-regulation implies that two important types of information are available: what precisely is expected of the worker and what precisely the worker is doing in response to the expectation.

An individual's immediate supervisor is more likely to be responsible for communicating expectations and specifying consequences. Greater clarification is likely to be created by periodic meetings between supervisor and workers and continual feedback, i.e., on-going monitoring of performance. Two types of information are needed: the results of the work and the personal relationships within the work or the "emotional and behavioural topography of the task".

The effective supervisor is expected to monitor work and make interventions as required. Failure to monitor makes it difficult to provide preventative supervision. Without monitoring, performance tends sooner or later to drift off track. For this reason, authorities see monitoring and feedback on performance as primary purposes of a good appraisal system. Monitoring needs to focus not simply on improvement of work but required changes in the work environment, better administration, and increase or change of staff or resources.

However, there are differences in the scope and effectiveness of supervision, depending on the organization. In business, the supervisor may carry authority for judgement and action and be directly accountable. In public organizations such as the educational system and, in particular, where bargaining units have a strong voice, there are often major restrictions on management, and this creates difficulties and ambiguities for educational supervisors.

A supervisor must also have sufficient expertise to be acceptable and do his work effectively. For example, in a scientific research program, the director may be quite unable to be expert in all aspects of the work and must ensure he can communicate with experts, providing them with appropriate monitoring and support.

Ondrack and Oliver confirm the point that the supervisor may have conflicting roles, as monitor and as judge. The authority-disciplinary role in appraisal can be distasteful to the supervisor as well as the appraisee. Further, the supervisor is often required to monitor, evaluate, and provide remediation, i.e., be both judge and coach. To cope with these demands requires effective training and acceptance of the role, possibly splitting the responsibilities between individuals or in a team in clearly defined ways.

Much is made in the literature (Ondrack and Oliver 1986, Ontario survey, Lawton et al. 1986) of the differences between formative and summative appraisal. The educational system, though acknowledging the need for monitoring and control of performance, tends to emphasize the claims of the formative appraisal and the purpose of professional improvement of the individual. However, the distinction is only relative and secondary. Purposes must depend on goals and circumstances. The distinction between formative and summative obscures the real purpose of appraisal to develop improved performance in the system. If monitoring and management of performance take their proper place, then the distinction between summative and formative is of secondary importance.

Performance appraisal is a means of telling an organization what it needs to know about its work, in many different ways, not just about the performance of individual workers.

The demand to establish formal appraisal systems occurs when the organization lacks confidence in established supervisory practices, i.e., in their fairness, consistency, efficiency in mobilizing resources, in motivating staff, etc. It is only in such situations that a formal appraisal system makes sense. It is not clear, in the Ontario context, whether the adoption of formal appraisal systems followed from the above lack of confidence or whether, as seems more likely, the formal appraisal provides a cover for the real and more fundamental processes of on-going supervision and monitoring. If, on the other hand, organizations are confident that supervisory practices give them the required information to monitor the system, an appraisal system may be used to direct worker improvement or to collect information more effectively for administration and planning.

Whether to use formal appraisal or not is an empty debate. Effective organizations do not debate whether or not work shall be monitored and controlled; they discuss the most effective ways of doing so. For the education system, the decision is not whether supervision and control of performance are desirable but whether a formal appraisal system is the most appropriate means.

Effective management must know staff and environment well enough to meet changing demands. Too much management is damaging, as is too little. Supervision must adjust to demand, whether in close, frequent involvement in some circumstances or encouraging independence in others. Despite reservations, an effective appraisal system may be the only concerted effort to ensure that workers (teachers, principals, administrators) know what is expected.

Performance appraisal and analysis of programs are closely related. The decision to implement a formal appraisal system should follow from assessment of organization needs. It asks questions such as: What information on the performance of educational staff is required by the organization to ensure that work is performed appropriately? How is the information gathered and used? Are current strategies for monitoring and control adequate?

Where supervision is already sufficient for effective care of the organization and individual performance, the development of a formal appraisal system is unnecessary. A school system may appropriately direct effort towards developing effective strategies for improving performance of some or all kinds or devote resources to provide information for administrative decisions. Even if the process for monitoring and control is deficient, the development of a performance appraisal system should have lower priority than development and implementation of more satisfactory supervision.

In summary, performance appraisal cannot be treated as a discrete activity, disconnected from other management. There must be broad understanding of the organizational context. The system must examine and evaluate its roles and responsibilities, have a clear sense of what results are to be expected from its work, and have standards of judging those results.

It is pointed out in the Hay Associates reports that knowledge of performance is needed to tell that the worker is on track. This depends on the nature of the work not simply on the worker. The Hay Associates reports develops the concept of three levels of work, based on the extent to which the worker is required to exercise discretion over results and the extent to which the worker has freedom to accomplish these ends by a variety of means.

Work in which both means and ends are externally set is a D-- pattern, that of the bus or streetcar driver, or factory-line worker. Work in which the objectives are fixed but means left to the worker is a D-+ pattern, the taxi driver or technician. Work in which both means and ends as determined by the worker is D++ pattern, typically the artist or professional.

The D-- pattern has standardized work processes and results, e.g., a bank teller, with little significant judgement needed. Most decisions for execution and production are made in advance. Production is controlled by schedules, by the book of rules, or by quality control. Management is passive, i.e., "no news is good news" and performance can be checked against required quality or the book. This has been called the machine bureaucracy.

The D-+ pattern allows the worker authority to make decisions but to be held accountable for results, so that the organization must have confidence, not only in the skills of the worker but the ability to exercise discretion on means. There is a focus on the selection of workers by skill, knowledge, and attitude. Where work assignments are on a brief time scale, control is by monitoring of final results to ensure targets are met. Where the work is part of a long time sequence or project, the worker will be expected to develop complex plans and submit these for monitoring to the management.

It is important to distinguish between work for which precedent exists and that for which there is no precedent and so provides a challenge in problem solving. Where the core work of an organization is concentrated on standardized results, the process will be better served if it directs itself to objectives rather than to challenges. For this reason, objectives should

not be set without clear ideas about how they can be achieved and providing the resources to achieve them.

Management by objectives can be a useful technique in the D-+ pattern of work. The semi-independent worker prepares work objectives and plans in concert with the supervisor, then is checked on achievement of the plan. This technique can fail and the worker become demoralized, if objectives are established without respect for their feasibility or the required skills or resources are not made available. This is obviously relevant in the teaching situation. A danger of the D-+ pattern is that it may emphasize results and have too little concern with how results are achieved.

The D++ pattern assumes that the worker, e.g., a social worker, doctor, lawyer, must discover the problem to be solved, i.e., undertake a diagnosis and only then decide how best to proceed. The response may depend heavily on the specific situation, e.g., circumstances, assessment of reasonable outcome, and the best means to achieve this. In this type of work, the role of the client is crucial.

The problem is to retain responsibility for performance when neither means nor ends can be specified ahead of time. One solution is to increase discretion of workers at the lower levels of a system which works in this manner, i.e., to encourage the "self-regulation" referred to earlier. In such a system, there must be a heavy emphasis on ensuring that the workers are rigorously selected for appropriate skills, knowledge, and professional expertise, since so much responsibility will be handed to them after appointment. There will be pressure to standardize skills, e.g., through university or professional training. Following the acquisition of formal knowledge and skills, there is often indoctrination into the particular demands, ways of behaving, and culture of the profession or organization, e.g., internship or probation.

Though the D++ pattern permits high discretion and choice in work, not any means or ends will do. The professionals of the organization must establish criteria and boundaries of accepted performance. In this situation, simply observing the worker's performance will not be sufficient, since the observer may not know with what assumptions the professional or artist is approaching the job, the diagnosis of the problem that has been made, and why particular means or techniques are selected as most appropriate. The observer may apply his own ideas of what is perceived as effective but this may not match with the worker's diagnosis and plans. Furthermore, D++ work does not occur in the most predictable or stable environments. The clients and the problems may present competing or conflicting demands or the demands and problems may shift from time to time. So it is necessary to modify expectations as well as procedures in appraising D++ work.

There are interesting, and in some ways paradoxical, implications for the appraisal of performance in different styles of work. In the D-- situation, it is enough to check on the match between actual and required production, i.e., the standard product or quality control. Feedback is specific and retraining directs itself to improving standards of production. In the D-+ situation, supervision is more complex and cannot be prescriptive. The supervisor may look for rough spots in the cycle of performance or the employee may ask the supervisor to look at specific areas where work can be improved. The supervisor of D-+ or D++ work must have expertise in the activity to be credible to the worker. This bears on the educational situation. How far, for example, can a principal be an expert in all aspects of what he appraises? How can a secondary school principal or a department head with ten years of administration and distance from the classroom evaluate specialized work in subject areas such as music, art, science?

If teacher and student performance can be measured and standardized, with the objectives set by administrators and the teaching done by the teachers, instruction is likely to be programmed and routine, i.e., a D-- situation. Then the monitoring and appraisal of performance can be carried out by direct observation. If, on the other hand, teachers are given discretion as long as they achieve predetermined goals, then the supervisor should be mainly concerned that the results are being achieved consistently. In this case, observation of the process in the classroom is not the appropriate appraisal.

If work at D++ professional level is being appraised, the appraisal must take into account the diagnosis, planning, choice of strategy and procedure, and appropriate self-evaluation of the work which has opportunities for choice and discretion. Once again, direct (classroom) observation is not appropriate. What is appropriate is discussion and agreement on relevant and effective goal setting, planning, and use of high-level skills, i.e., analysis of the complex process of professional work but also its outcomes. In no interpretation of teaching as a high-level professional activity is there much place for reliance on a narrow focus of classroom observation as a sole or main feature of appraisal.

It is clear, therefore, from the literature review and the Hay Associates report argument, that the nature of the task must be determined before criteria, standards, and information can be properly brought to bear in appraisal. It is equally clear that there is not an acceptable theoretical or practical analysis of what is involved in teaching. Review of the kinds and levels of tasks required in teaching suggest that it should fall within the D-+ or D++ work patterns.

There are four possible forms of work organization. The first, professional bureaucracy, characterizes many professional organizations such as social work, accounting, research organizations, and medicine, where the professionals operate within an organization or institution, not entirely on their own as entrepreneurs. Professional work of this kind has a high level of autonomy and decision making. The high level of autonomy is claimed because of the high level of professional training and selection of staff. A profession requires, not

only a body of specialized techniques, but judgement about the application of these and their relationship to a body of theoretical knowledge, e.g., medicine, the law. The crucial question is whether teaching, in its claim to be a profession, has these characteristics. In a professional bureaucracy, there are constraints imposed by the working context and the relationship between professional and client. Outputs may be difficult to measure directly. Much may depend upon agreement between independent professionals on judgements of the quality of their work.

One very important and possibly disturbing implication for appraisal is that managers can do little to correct deficiencies directly if professionals fail. Discretion not only gives professionals freedom to do their best work, but can enable them to ignore the needs of the organization. There is a danger in professional work of "means-ends inversion", in which workers become so convinced of their skills and satisfied in performing them that the needs of the client may be overlooked as the professional follows out his favourite program. One may ask whether teaching, with its emphasis on process and the independence of the teaching act is subject to this danger.

The greatest barrier to acceptance of work as fully professional (D++ pattern) is the lack of a body of specialized theoretical and empirical knowledge and techniques which is fully accepted by the community. Is the educational system structured and managed to support a D++ conception of teaching? Is the primary role of supervisors to protect teachers' autonomy and ensure they have educational resources? Are teachers, like true professionals, directly accountable to their clients or to the clients of the organization? These are all questions which arise from consideration of the nature of professional work and appraisal.

Within the educational system there is a wide variety of stakeholders requiring accountability. This reduces the range and autonomy of the workers in this particular professional bureaucratic system. The social and political accountability of the school system points to the need for a degree of centralized monitoring and control to achieve negotiated goals. This reduces the scope and applicability of the full professional bureaucracy model to teaching.

Another model, in organization theory, is the divisionalized form. This is the creation within an organization of specific subdivisions devoted to specific goals and outputs, e.g., the different divisions of a motor industry. There is delegation of authority and control to the divisional units, which may be loosely or closely coupled. There are some parallels in education, in the division between age and grade levels and functions of schools and the specialization of roles at the secondary school level, but the divisionalized form is not a close model for education. This model requires supervision to ensure control and standardization of overall production while allowing for flexibility, e.g., possibly like the departments within community colleges. The model is most effective when there is a single consistent set of final goals in the form of output, sales, profit, etc. This is why this form appears more relevant at the community college level where there are discernible courses

with specific vocationals, aims, or credits. The divisional form does not function well outside industry, in spite of attempts to adopt it in universities, hospitals, and government. Without standardization of skills and measurement of output, the system does not work well.

Another possible organizational model for teaching is the rationalistic model. Professionals in this setting perform D-+ or even D-- tasks. Individual discretion in dealing with a client is subordinated to the task, organization, and co-workers. This is similar in many respects to teaching, which is only a partly independent set of activities within a larger educational context, with many restraints. As the educational system becomes larger and more complex, following the rationalistic model, the administrators become increasingly important in organizing the system. The model requires: functional division of labour; definition of staff roles by function and performance; hierarchical ordering of office; and operation within procedural rules which limits discretion on aims and procedures. In this model (D-+ or even D--), a teacher is subject to some prescription and standardization to foster co-ordination of effort and is responsible for specific results, while management remains responsible for accountability to the wider community. Appraisal may combine direct supervision, mutual negotiation of goals and procedures, prescription of work, or standardizing procedures. In education, standardizing may include curricula, grade levels, credits, etc. This may be a better fit to a number of aspects of teaching, though not entirely flattering to the full professionalism of teachers.

Finally, a "natural systems" model of organization has been proposed. The characteristics of this system are: absence of consensus on values, norms, and objectives; functional autonomy of parts of the system; bargaining and compromise on goals, plans, and procedures; de facto centralization of power and authority; incomplete information for making decisions; and lack of co-ordination in planning and policy making. These are somewhat disconcerting features of a working organization, but the question is how many of these appear in the present educational system? There are implications in the model for the organization and for performance and program appraisal.

If there is a long hierarchy between classroom and higher levels of work, the higher levels tend to relate to the outside social and political system and be susceptible to its pressures. There is possibility of misinterpretation of activities, goals, and plans, and conflict of interest between different levels and aspects of the system because it is so loosely coupled. Because functioning is decentralized (down to area, school, and possibly even classroom), policy and planning are not highly co-ordinated. Teachers may not govern their routine activities by reference to deliberate and planned decisions within the educational system. The strong influence of bargain and compromise means that it may not be possible to override the individual's conception of what the work is. However, the considerable discretion exercised by teachers is by default, unplanned, and not by choice, i.e., the teacher may appear to be operating as a high-level (D++) professional, but is not.

Monitoring and appraisal based on prescription of work processes (methods) or of results are both of limited value. The primary co-ordination must therefore be an insistence on the standardization of the skills, knowledge, and effectiveness of the worker (teacher, principal, superintendent, consultant). This, in turn, means that one of the few effective ways of ensuring adequate management, monitoring, and control of work is through selecting only highly trained and skilled workers and then ensuring that they are indoctrinated with the skills, procedures, and values of the system. This is, in fact, what occurred in the Ondrack and Oliver site studies of American schools, where some school systems, in order to ensure quality of education and control of the management-appraisal process, placed heavy emphasis either on certification and probation or on training staff in the specific methodology adopted by the school system. Does the Ontario school system devote this degree of attention to the crucial stages of selection, certification, and probation of new staff?

A devastating implication of the natural systems model, with high task-uncertainty and looseness of organization and accountability, is that it operates by setting low performance standards which are just sufficiently viable to avoid a breakdown of confidence among the clientele or those funding the system. In this context, inspection and appraisal are designed to avoid highlighting inconsistencies and ineffectiveness of performance and maintain the appearance of effective performance. The absence of consensus on values, objectives, and procedures in the natural system is a serious obstacle to the development of performance criteria or standards. The existence of a natural system or even substantial elements of it, would raise fundamental questions about the meaningfulness of accountability and performance appraisal. Is the educational system in Ontario such a system? What crucial and disturbing questions are raised about the teaching profession?

The Hay Associates reports points out, possibly rather cynically, that it may be said that managers of educational enterprises must not only manage performance, but also, even more important, appear to be managing performance.

Educational systems, it appears, cannot guarantee a close coupling of activities of instruction and outcome. They depend on a logic of confidence in which each successive level of the system organizes its work on the assumption that what goes on at lower levels makes sense and is producing appropriate outcomes. In this way, the system can, at any given level, avoid inspecting itself or assuming overall responsibility for inconsistency of effort or for effectiveness. Planning and policy at the board level, for example, may make assumptions about appropriateness of curriculum, staffing, resources, and performance at classroom level but with only a loose coupling between the political and policy level and the classroom outcome.

In setting up performance appraisal systems, there are many snags. For instance, any attempt to increase public confidence in education simply by rigorous implementation of appraisal will backfire unless teachers have appropriate qualifications, resources to meet the plans are provided, and policies are clearly implemented. Performance appraisal systems

create an opportunity to subject the bases of educational confidence to scrutiny. The bottom line is the appropriateness of performance compared with expectation.

Much of the concern in the past decades in developing appraisal systems has been to deal with possible problems or weaknesses of supervision and monitoring of the system by attempting to replace these with valid and reliable judgements of performance, with specific, defensible criteria. Effective organizations, however, establish their standards of performance long before adopting an appraisal system, if they ever do. Standards of performance must deal with basic elements of job design, the reason for the work, who is to work, and the work processes to be managed.

What is crucial, therefore, is that an organization must, before getting involved in performance appraisal, define its functions. It must ask basic questions such as: (1) Has it failed to identify the needs of its clients and is either providing services no longer required or not providing what is now required (e.g., program, organization, resources)? (2) Is it failing to meet established standards of quality for its services? (3) Has it correctly identified what it should be providing, but has not set up standards of quality?

Performance appraisal may be an inappropriate or inefficient solution to a given problem as compared with other direct changes that may be made which may be less costly. If the organization is providing services no longer appropriate to the needs of clients or failing to provide new ones, clearly the organization must review its functions and adjust the needs-to-service contract. If the organization is failing to meet established standards, this may be due to inadequate resources, poor co-ordination of effort, insufficient budget, or inappropriate training or selection of staff, not necessarily failure in individual performance. If the organization has failed to set performance standards, it should establish what these standards are, determine if it has the appropriate resources and expertise, and take steps to ensure that all staff have the required skills, knowledge, and attitudes. That is, management must identify competence but then ensure that this competence can be generated and will meet the required standards of performance.

The implication of this approach is that educational organizations must undertake comprehensive analyses of their functioning and program. The system must review the primary and in-service training of teachers and ensure effective selection and induction of staff at the certification, appointment, and probationary levels. There is no evidence that the introduction of performance appraisal systems in Ontario was based on such fundamental analyses of the needs and goals of the educational organization.

It is crucial to distinguish between situations where there are established standards and where there are not. If a school system has a model of teaching effectiveness and has trained teachers to meet those requirements, it can appraise performance. But if there are no standards, it is inappropriate to import them wholesale. Bringing in new standards and competences, and appraising performance against these standards is dysfunctional to the

organization and unacceptable to the workers. Accountability is less contentious if the persons appraised have had sufficient opportunity to understand the standards and acquire the required skills.

For any review of teaching and appraisal, educators need to ask: (1) Do we know enough about teaching and learning to justify use of performance standards? (2) Do we know what competences are required to meet standards? (3) Have teachers been given adequate opportunity to develop competences? If the answer is yes to the first and second questions, then the issue of adequate training and provision of resources is the important one.

In this connection, it is pointed out by the Hay Associates report, as also indicated by the Ontario survey (Lawton et al. 1986), that a high proportion of teachers, especially at the secondary level, were trained well over a decade ago before present demands for competence and performance were called for. Unless school boards have carefully and systematically retrained and indoctrinated teachers in required new competences, there cannot be justice in accountability, i.e., the rules are changed after the play begins. The emphasis on seniority, and consequent aging and obsolescence of the teaching force, has serious implications for performance and appraisal.

In looking at the organization and appraisal of work, there are three criteria: (1) relating to the worker, such as skills, knowledge, attitudes, and qualifications, (presage criteria); (2) relating to work, such as practices, procedures, operations, plans, and behaviour (process criteria); and (3) relating to the results of work, such as services, attainment of students, changes in student learning (outcome criteria).

If one looks at the patterns of work, it is clear that process measurements are most appropriate at the D-- level, i.e., monitoring that the work is correctly done. Results are the most appropriate measure for D-+ work where the teacher can choose means but not goals. For D++ fully professional type of work, it is crucial to use presage or qualification criteria in addition to the others, to make sure that the person who is going to work autonomously is appropriately qualified and skilled.

It is clear that the observation of classroom performance, in itself a narrowly focused criterion of performance and the method in most educational appraisal, is, in fact, appropriate only to relatively low levels of professional work. The use of presage criteria, i.e., emphasis on the training, qualifications, and professional skills of workers, with the focus on accurate selection and probation of professional staff, implies that there is explicit understanding of what constitutes professional preparation and qualification and what teaching is.

Less is known about this than some believe but more than pessimists believe. But even if it is known what effective teaching is, knowing the competences required is another matter. In turn, even if effective teaching behaviour is defined, and competences measuring and

predicting this behaviour are discovered, one has still to ensure that these relate to actual outcomes in terms of student learning. Generic lists of competences, checklists of observed classroom performance, tests of teachers' ability or knowledge or other characteristics may have only little or an indirect relationship to actual student achievement. And in the last analysis, student learning and achievement are the criterion for effectiveness in the educational system.

Screening and selection of staff does not, however, eliminate the need for continued monitoring of performance.

The most appropriate way to control performance in a professional setting is to make the appraisal self-regulating, i.e., by providing information to the worker on what is expected by the system and how the work is going. Greater clarification of both performance standards and role responsibilities in an organization is provided by regular meetings and feedback where the work objectives are established. This has serious implications for formal appraisal systems where teaching staff are formally appraised and given feedback on performance only once in a number of years.

A worker bases judgements on information not only from a supervisor but from colleagues, subordinates, and clients (students), i.e., both internal and external information. This feedback is crucial. This has implications for educational appraisal systems (American field studies, Ontario survey) which confine themselves to a narrow range of criteria such as classroom observations. Clearly, a greater variety of sources of information on teacher performance is needed.

Ability to distinguish relevant feedback is a function of how the worker understands and accepts what is expected. The control process requires that organizational goals should be met at both the individual and school level.

The minimum assessment of highly professional (D++) performance requires an assessment of the goals and standards of the worker, to ensure that they are consistent with those of the organization. In the last analysis, D++ work is most efficiently monitored by attention to results rather than processes. However, D++ professionals should provide their own yardsticks for effectiveness, as contrasted with the D-+ or D-- worker who has limits on discretion and choice. For these reasons, the distinction between formative and summative evaluation, so much emphasized in educational appraisal, is secondary and obscures the real purpose of appraisal. Action to stimulate improvement or to change employee status needs to be based on well-founded supervisory judgement on how the needs of the organization are being met.

It is clear that this discussion points to the primary importance of effective management, supervision, and monitoring. Only when the organization lacks confidence in established supervisory practices does the use of a formal appraisal system make sense.

The issue of training of teachers to meet the requirements for performance is crucial. It is clear from the Atkinson report that expert educational opinion in Ontario considers present teacher training programs less than adequate for professional preparation. The educational system, in dealing with the possible problems of insufficient primary-stage professional training might claim the compensating effectiveness of in-service and specialist training. However, as the Ontario survey (Lawton et al. 1986) and the Ondrack and Oliver field studies 1986 demonstrate, there is no guarantee that individualized professional improvements are available, required, or followed up, in spite of the claims that appraisal deals primarily with professional development.

The present situation in professional training implies that teachers will continue to work on their own, i.e., have a self-managed "internship" of indeterminate length. It will never be quite clear when a teacher is professionally qualified. This is contrary to all other professional training which specifies the length and nature of professional training, including internship or probation.

If there is such a vague limit to the definition of professional training and performance, two fundamental questions arise: What is the basis for applying external and post hoc standards to monitor teacher performance which is so variable and idiosyncratic in terms of training and preparation? At what point is it appropriate to treat a teacher as fully qualified?

This is one reason why educators assert that appraisal should be formative, since a formative evaluation avoids the issue of specifying goals, practices, and precise performance and substitutes a general concept of improvement over present status. This, however, is a substitute for effective supervision and appraisal.

Clinical supervision is one possible model of professional training, i.e., the close association of a professional with a colleague or supervisor in which progress, practices, and professional learning are periodically reviewed in the context of the work setting and professional practice, with frequent feedback leading to modification of plans and procedures as required. This is a truly formative professional approach. However, the technique is complex and time consuming.

Much emphasis has been placed on process appraisal or goal-free evaluation. The most compelling reason for this focus may be that alternatives such as the use of results (e.g., student learning) are perceived as impractical or impolitic. The question, however, is whether the difficulties in using outcomes are insuperable or have been exaggerated.

The use of student achievement as a basic criterion of teaching effectiveness has four arguments advanced against it.

1. It is claimed that the impact of teachers on student achievement is limited.
2. A host of conditions beyond the teacher's control have a major influence, e.g., number of students in the group, student ability and prior learning, time, curriculum materials and methods, and classroom organization.
3. The conventional achievement tests and examinations tend to reflect limited aspects of student performance, emphasizing skills and rote information as contrasted with complex learning.
4. When evaluated on test or examination performance, teachers tend to teach to the test.

The above provide a pessimistic view of the effectiveness of teaching. If really believed, it would suggest that there is little purpose in considering teacher performance, which would place the teacher in a very strange position indeed.

As noted earlier, these considerations point to the need for effective program evaluation in its broadest sense, and the use of innovative procedures for testing and calibrating student learning. There are examples of the effective use of student achievement within a broader appraisal framework, as reported by the Montgomery County, Maryland school system (Ondrack and Oliver field studies 1986) and demonstration by research studies such as those of Ryan and Hickcox (1980) reported in the Ondrack and Oliver literature review, that outcomes such as student achievement can be built into an appraisal system which is consistent with a view of the teacher as a professional.

In summary, appraisal must always have a prescriptive, monitoring function and cannot depend solely on process appraisal. Judgement on professional (D++) performance must incorporate information on all aspects of performance, i.e., diagnosis, learning, goal setting, strategies, and execution, as well as outcomes in student learning. Appraisals may need to be tailored to individual styles and different states of career and experience.

Being multilevel is both the greatest strength and greatest weakness of an appraisal system. It seems unlikely that all the various needs of appraisal and monitoring, ranging from certification and probation to promotion and special cases can be met by the same system or same approaches. A supervisory or appraisal system must have different approaches and techniques for its various purposes. For example, there will be particular difficulties in reconciling an appraisal which deals with job status or compensation with one emphasizing professional improvement.

To repeat and summarize: the organization must know what performances it wants and what information it requires, and this will derive from the view held of the nature of professional work and, specifically, of teaching.

Quite clearly, an efficient supervisory and monitoring system, with consistent and frequent feedback to those appraised and with appropriate and immediate forms of action to change and remediate, is to be preferred to a formal appraisal system. Where supervision

already provides sufficient information for control of organizational and individual performance, formal appraisal is unnecessary and inefficient. Energy would be better directed to strategies for improving organizational and individual performance, e.g., teacher training, revision of curriculum and program, altered forms of school organization, improved resources, and improved administrative information and action.

The consequences of effective performance monitoring in an organization are for change, as follows: allocation of resources, redesigning of tasks, adjustment of standards, redistribution of workload, and training or remediation for the individual.

The ability to take appropriate action is based on the expertise of the supervisor, as well as the restrictions imposed by the organization, including special interest and political and social pressures. If supervisors cannot provide the expertise directly, then they must be provided with the required advice and support, e.g., a principal may need administrative advice, or consultation in specific areas of the curriculum. The clearer the expectations of task and of role, the more effective supervision can be.

A useful distinction may be made between work which is of higher or lower order of complexity or which has a higher or lower degree of uncertainty in terms of what the tasks is, what appropriate goals are, and what are the best plans for accomplishing the task. The D-- work pattern is likely to be of lower complexity, predictable, consistent, and programmed. Progress can be inspected and rewards related directly to results, with other incentives having little effect. The D-+ pattern, with no discretion as to goals, may respond to incentives to the individual and group, e.g., pay for performance or equivalent forms of motivation. At D++ professional level, work is not effectively dealt with by regulation, formal authority or external awards, but by satisfaction in completing the task and internalizing the goals and values of the profession on the one hand and of the organization on the other. The work is likely to be more complex and more uncertain in many respects. If attainable goals are clearly specified, however, and the D++ individual identifies with them, there is greater likelihood of co-operative behaviour. This depends on well-defined professional norms and standards. Generalized, vague goals may apparently allow more freedom, but at the expense of effective cooperation. The natural system model of work illustrates these problems.

One of the crucial issues in supervision or performance appraisal is that of dealing with inefficient or unacceptable performance and undertaking remediation and review. Most organizations take this question seriously (Ondrack and Oliver 1986) not only because of the serious implications of incompetence for the individual, but also because of the potential administrative and legal difficulties involved in grievance and dismissal procedures. Most organizations therefore provide, even if not in the formal appraisal system, for detecting consistently substandard performance. The appraisal system needs to identify the obstacles to performance and whether the individual or group is demoralizing clients or other colleagues by being incompetent or uncooperative.

The need for fair, valid, and reliable appraisal at this stage is crucial. Efforts would normally be made to remediate the problem, with additional support or retraining, or to reallocate the person, with a time line for the salvage operation before final action is taken to dismiss. In practice, educational systems may transfer teachers or redesignate expectations and duties to avoid dismissal. As noted in the Ontario survey (Lawton et al.) the actual proportion of teachers or principals coming to grievance or dismissal is remarkably small, arguing either for the very efficient operation of appraisal or the converse.

At the D-- work level, the principal's direct observation of substandard performance might lead to proposals for change in practice, e.g., additional training or remediation. At the D-+ work level, regular monitoring of performance would be unlikely to uncover the sources of the problem, as actual areas of performance and outcomes would be in question. The obvious improvement required would be in a raised level of student achievement, but the teacher could not be held exclusively responsible for this improvement, as factors such as student ability, organization, program, etc., would need to be reviewed and possibly changed. For the D++ professional, efforts would likely be aimed at improving specific aspects of skill or knowledge felt to be below standard, e.g., in diagnosis or in setting objectives. Where a D++ worker is placed on review, the total effectiveness of the individual is in question. At this point, observation by a skilled observer or appraisal team would be in order, to deal with both process and outcome and the total pattern of performance.

Clearly, there must be stringent requirements governing special cases and review which could lead to dismissal. There must be four requirements:

1. The employee must be informed of the standards on which he/she is appraised.
2. The employee must be notified of the serious deficiencies to be corrected.
3. The employee must be given the opportunity and assistance to improve.
4. There must be reasonable time for correction, remediation, and salvage.

There must also be a clear distinction between the need for and provision of professional development as such and the specific remediation of incompetent performance. This is by no means clear in formal appraisal systems such as those described in the Ontario survey (Lawton et al. 1984). Failure to make this distinction confuses the whole issue of effective appraisal.

The review-appraisal should also distinguish between improving competent performance on recent standards and the need to adjust to new demands and standards in the organization (e.g., new program, materials, methods, or forms of organization). When organizations change their goals or programs, as may follow from a comprehensive review which should always precede the adoption of appraisal, then teachers must be informed that they are going to be expected to perform to different standards and be given opportunity to develop the required competences. This is a crucial issue, bearing in mind the large number of teachers without consistent updating of professional skills and knowledge and the current reliance on seniority.

This discussion points to the need to have an efficient system of primary professional training, certification, and probation if teaching is to be regarded as a D++ level profession. Equally, it points to the need for a comprehensive, organized system of individualized in-service training for completion of professional status, as well as for advanced professional development and special cases such as remediation.

It should be recalled that all established professions have inbuilt measures for up-dating and retraining their members, some including the requirement that only if members show evidence of such up-dating may they continue to hold professional status (e.g., in the field of marital/family therapy). The teaching profession needs to take this requirement seriously if it is to be regarded as fully professional or self-regulating.

Where standards or program change, as they must, the organization has a clear responsibility for ensuring professional development and adjustment. The organization must make it possible for the individual to find the appropriate help and ensure that it is followed through, unlike the present situation in appraisal in the United States and Ontario.

Appraisal directly related to compensation (payment) benefits are difficult to integrate with appraisal which attempts to ensure organizational goals are met and personal professional development ensured. Compensation should not be a major factor for the D++ professional worker. This raises the difficult issue of additional payments for scholarship as distinct from rewards for acquiring additional specialist qualifications. Presumably, such qualifications and rewards should be tied by the organization to specific uses and functions rather than to the person, e.g., a teacher should not claim specialist qualifications as meriting continued compensation in contract negotiations when these qualifications are no longer actively needed because of change of circumstances. On the other hand, there would seem to be a case for offering incentives to proceed to higher professional qualifications, such as becoming a master teacher or acquiring advanced professional (as contrasted with academic) qualifications. As noted in the discussion of the American school system field studies (Ondrack and Oliver 1986), the formal appraisal system appears to be ambivalent and lacking in purpose in dealing with this kind of incentive. Once again, this is a matter for effective management and supervisory practices rather than formal appraisal.

Appraisal of professional development includes: collection of information on performance on the job; identification of performance strengths and weaknesses; articulation of objectives for performance improvement; identification of resources needed; monitoring of progress towards improvement; and final re-evaluation of improvement.

The three main purposes of appraisal are: monitoring the performance of employees against organization goals and standards; stimulating professional development in competent employees; and remediation of substandard performance if necessary.

As previously discussed, one of the most relevant models of appraisal and development for the professional is clinical supervision, which can take place in collegial relationship between supervisor/colleague and professional workers. This is more appropriate than a formal system of performance appraisal. However, there are practical difficulties of complexity, training, and time, and a crucial point is whether teaching is sufficiently and consistently enough at the professional level to make this practice feasible and acceptable.

The following summarizes some of the major arguments in this section. The constant argument is for specific and flexible forms of supervision and monitoring, with the use of appraisal for specific defined purposes only when justified and required. Again and again, the need to place performance appraisal in the context of the organization's purposes and goals has been raised. The fundamental question is, What is the business of the organization and what should it be? Performance appraisal is not a panacea. In many respects, it has not been well done. Many questions must be asked before a formal appraisal system is adopted, if it is adopted. Solutions to the problems encountered by the organization may well be solved in other and better ways than by undertaking individual performance appraisal. As long as management and supervision are effective, there is no need for formal appraisal.

Fundamental questions need to be answered about the nature of professional work and the form of organization in which professional work can take place, before performance appraisal can be meaningful. The discussion of the D-- to D++ models of work and of the various possible models for teaching raises serious questions and distinct doubts about whether teaching is entirely a professional activity in these terms.

If the educational system is best described by the natural system of loose, unorganized relationships and responsibilities, as seems plausible, there are fundamental questions about teaching, accountability, and confidence in the educational system. Fundamental questions are raised about the purpose of process and formative evaluation and professional development as a basic aim, unless seen in the context of organizational goals and needs. The need to consider outcomes and results as fundamental components in appraisal of teacher performance and the need to overcome problems in making use of outcome data is also reviewed.

Crucial questions are raised about the effectiveness of teacher training and in-service training as preparation for professional status, the lack of definition of when professional status is complete, and whose responsibility this is.

There are clear implications for the importance of program evaluation in its broadest sense, as prior to performance appraisal, as a necessary implement to it, and as the necessary instrument and sufficient technique for assessing the effectiveness and validity of performance appraisal.

Conclusions

1. The present set of studies on the performance appraisal of educators has assembled a formidable amount and range of information based on a comprehensive literature review and a well-organized and representative group of field studies which enable comparative judgements to be made of appraisal in important non-educational professional organizations, within different forms of education, and between the United States and Ontario.

A large-scale, comprehensive review of performance appraisal of educators at all levels in the school boards of Ontario produced a representative survey of the state of the field in the 1980s.

A review of evidence, principles, and problems underlying the use of performance appraisal in organizations was presented, enabling the reader to focus on major conceptual and organizational issues. In many respects, this review by Hay Associates is the single most useful section of the total research effort.

2. A considerable benefit of the literature review on performance appraisal was that it was conducted by experts in management and organization studies and therefore related the issues in appraisal in education to the broader issues and principles of the function and history of performance appraisal in a wide variety of organizational settings including industry, business, and research. It thus removed discussion of appraisal from what might be termed a narrow psychological and educational perspective and placed it in the perspective of organizational theory.

3. The large-scale empirical study of performance appraisal in the school systems of Ontario assembled a considerable amount of basic information on formal appraisal systems and the views of those involved in them. However, the impact and value of the survey was limited by its educational focus, with limited acknowledgement of the organizational background and its dependence on questionnaire data and report. The emphasis on process meant that the question of ultimate criteria such as the effect on student learning was not raised. The study dealt only with formal appraisal practices and could not examine management and supervisory practices, though a substantial minority of school boards in Ontario do not have formal appraisal systems.

4. The summary of the empirical study indicated that there is a plethora of appraisal systems in Ontario at present, with considerable effort devoted to designing and running these systems, but that the effort spent on developing and implementing appraisal is hardly commensurate with the observed benefits. To say the least: "Some thoughtful changes will have to be made in most board policies and practices".

5. In terms of research, therefore, there is no purpose in pursuing further research with similar techniques on formal performance appraisal systems in Ontario.

6. Considerable effort has been devoted to designing and establishing performance appraisal systems in Ontario but there is little commitment to evaluation of their effectiveness and maintenance. Boards need to evaluate the effectiveness of their appraisal systems and appraisal systems need to be maintained, improved, or abolished depending on evaluation and feedback from all involved.

7. In seeking to establish or improve performance appraisal systems, boards of education cannot assume there is "one best model" or simply copy practices or instruments, but need to develop a system which best meets the purposes, goals, and circumstances of their organization.

8. Boards of education need to review current evidence and make use of it in improving appraisal and supervisory practices. In particular, the definition and description of appropriate teacher competences should be considered.

9. The current studies show that there is wide local variation in appraisal systems.

10. The current studies demonstrate clearly that performance appraisal systems must meet organizational needs and goals and that formal appraisal systems should be adopted only if management and supervisory practices are ineffective. The evidence shows that a substantial majority of boards of education in Ontario carry out their management and supervisory functions without formal appraisal systems.

11. The current studies make clear that appraisal and supervision must take place in the context of the needs, purposes, goals, and problems of the educational organization. Any appraisal system must be developed in relation to the needs of the organization, and only if specific needs exist or problems exist which cannot be better dealt with by other means.

12. The current evidence is that the evaluation of the effectiveness of performance appraisal is best served by program evaluation. In general principle, program evaluation is the mirror image and complement of performance appraisal. Effective teaching is demonstrated by appropriate and effective use of curricular and program materials and methods. Conversely, curricular materials and methods or programs are evaluated through effective application in teaching and in student learning. For these reasons, research and development in program evaluation play a crucial role.

13. Serious consideration needs to be given to the use of outcome measures (results) in the appraisal of professional work in education. Such outcomes are essential in program evaluation. The evidence from analysis of professional work in the present studies indicates that the assessment of student learning could contribute significantly to performance appraisal. Attempts have already been made (e.g., Montgomery County, Maryland) to use student achievement as part of a system of accountability. The assessment of student achievement, program evaluation, and performance appraisal of the teacher/educator are all interrelated.

14. It is evident that focus on classroom observation is not an appropriate form of appraisal if teaching is a fully professional activity, but there have to be wider and more varied forms of appraisal, such as the assessment of diagnosis, planning, choice of strategy, choice of materials, and methods in teaching. Techniques of appraisal such as self-appraisal, peer appraisal, the use of student evaluations or feedback, and management by objectives are all mentioned in the survey of Ontario practice, but used in only a small minority of cases or not favoured by teachers. However, there seems to be scope for more use of or experiment with these techniques.

15. The findings of the present study raise fundamental questions about the effectiveness of the present provision of teacher training for full professional development. The stage of certification, probation, and first entry to profession in the internship stage is highly critical for professional status and quality and needs special attention in appraisal and supervision, as field studies of American schools have shown.

16. Critical questions have been raised by this study about the effectiveness of professional development. It is evident that, though lip-service is paid to it, little provision is made for the individualized professional development to improve performance and less for follow-up and monitoring. It is evident that the whole system of in-service training and professional development requires an overhaul, if professional development is to be planned and individualized for educators at all levels of the educational system.

17. Urgent attention is needed for developing appropriate appraisal and supervisory techniques for monitoring the work of senior educational administrators and for developing training courses in appraisal at this level of the educational system.

18. The educational system has to demonstrate, through formal appraisal or other supervisory approaches, that it has a valid, just method of detecting substandard performance and problems at an early stage and of instituting remedial efforts or proceeding systematically to review and, if necessary, to dismiss. It is concluded in the review of evidence that incompetent performance is a burden on the whole system.

19. It would be beneficial to develop the technique of the appraisal team, at present used in a minority of situations in Ontario but positively received and it seems there is potential for developing this technique.

20. It would appear that research and development in the field of educational evaluation must take account of wider perspectives in organization and management studies.

References

1. Shulman, L.S. "Those Who Understand: Knowledge Growth in Teaching", Educational Researcher 15 (February 1986).

2. Hoffman, J.V., ed. Effective Teaching of Reading: Research and Practice, chapters 1 and 13. Newark: International Reading Association, 1986.